

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:04:37 ; Search time 11.3797 Seconds  
(without alignments)  
273.300 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALEALADQTLQSEEAHVVKADNAA 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	129	100.0	28	21	AA1980
2	129	100.0	141	21	AA1981
3	129	100.0	155	21	AA1982
4	129	100.0	159	21	AA1983
5	129	100.0	167	21	AA1984
6	129	100.0	184	21	AA1985
7	129	100.0	217	21	AA1986
8	129	100.0	259	21	AA1987
9	129	100.0	271	21	AA1988
10	57	44.2	192	18	AA1989
11	57	44.2	222	18	AA1990

12	57	44.2	412	19	AA1994
13	50	38.8	1164	22	ABG30268
14	48.5	37.6	450	18	AAW11567
15	47	36.4	258	20	AA136771
16	46	35.7	705	22	AA190301
17	45.5	35.3	73	22	AB143356
18	45.5	35.3	73	22	AA14267
19	45.5	35.3	73	22	AA17089
20	45.5	35.3	73	22	AA17212
21	45.5	35.3	521	21	AA1518
22	45.5	35.3	572	21	AA1517
23	45.5	35.3	2076	22	AA14319
24	45.5	35.3	2186	22	AA17320
25	45	34.9	476	22	AA13617
26	45	34.9	502	21	AA15480
27	45	34.9	1362	20	AA199481
28	45	34.9	1447	20	AA181029
29	45	34.9	1637	22	AB160342
30	44.5	34.5	515	19	AA161907
31	44.5	34.5	516	19	AA170356
32	44.5	34.5	527	19	AA170357
33	44.5	34.5	552	19	AA161908
34	44	34.1	42	22	AB101077
35	44	34.1	274	21	AA158293
36	44	34.1	274	21	AA161028
37	44	34.1	284	22	AA178512
38	44	34.1	296	21	AA158292
39	44	34.1	296	21	AA161027
40	44	34.1	344	19	AA161234
41	44	34.1	358	21	AA158291
42	44	34.1	358	21	AA161026
43	44	34.1	371	21	AA181672
44	44	34.1	421	22	AB171817
45	44	34.1	429	22	AB165976

#### ALIGNMENTS

#### RESULT 1

AA171038

ID AA171038 standard; peptide; 28 AA.

AC AA171038;

XX AA171038;

XX 29-AUG-2000 (first entry)

XX Streptococcus pyogenes strain SF370 GRAB protein fragment #3.

DE GRAB protein; protein G related alpha2M binding protein; vaccine;

XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

KW immune response; Streptococcus pyogenes infection.

KW Streptococcus pyogenes.

XX WO200026240-A2.

XX 11-MAY-2000.

XX 02-NOV-1999; 99WO-GB03631.

XX 02-NOV-1998; 98GB-0023975.

XX (ACTI-) ACTINOVA LTD.

XX Björck LH, Rasmussen M;

XX WPI; 2000-365572/31.

XX New alpha2M binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein

XX Claim 3; Page 56; 67pp; English.

PS

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is one of the repeat regions of GRAB protein from  
CC S. pyogenes strain SF370 corresponding to residues 92-119. This fragment  
CC is useful in vaccine composition.  
XX  
SQ Sequence 28 AA;

Query Match 100.0%; Score 129; DB 21; Length 28;  
Best Local Similarity 100.0%; Pred. No. 1.2e-11;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SDALEALADQTDALQSEEAAYVVKADNAA 28  
|||||  
DB 1 SDALEALADQTDALQSEEAAYVVKADNAA 28

RESULT 2  
AAV71040  
ID AAV71040 standard; peptide; 141 AA.  
XX  
AC AAV71040;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 GRAB protein fragment #4.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX

XX New alpha2M binding protein for generating a protective immune response  
CC to group A streptococcus and purifying the binding protein -  
XX  
XX Claim 5; Page 57; 67pp; English.

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-174. This fragment is devoid of the trans-membrane  
CC and cell wall anchor regions. It is useful in vaccine composition.  
XX  
SQ Sequence 141 AA;

Query Match 100.0%; Score 129; DB 21; Length 141;  
Best Local Similarity 100.0%; Pred. No. 7.6e-11;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SDALEALADQTDALQSEEAAYVVKADNAA 28  
|||||  
DB 59 SDALEALADQTDALQSEEAAYVVKADNAA 86

RESULT 3  
AAV71044  
ID AAV71044 standard; Protein; 155 AA.  
XX  
AC AAV71044;  
XX

XX 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain AP1 partial GRAB protein.  
XX

KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX

PH Key Location/Qualifiers  
FT Misc-difference 17  
FT /note= "Thr at position 18 of GRAB protein fragment  
FT (AAV71036) from S. pyogenes strain SF370 is replaced  
FT with Ile"  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX  
PR N-PSDB; AAD00562.  
XX

PT New alpha2M binding protein for generating a protective immune response  
CC to group A streptococcus and purifying the binding protein -  
XX  
XX Claim 5; Page 60-61; 67pp; English.

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a partial GRAB protein from S. pyogenes strain AP1.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.  
XX  
SQ Sequence 155 AA;

Query Match 100.0%; Score 129; DB 21; Length 155;  
Best Local Similarity 100.0%; Pred. No. 8.4e-11;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SDALEALADQTDALQSEEAAYVVKADNAA 28  
|||||  
DB 58 SDALEALADQTDALQSEEAAYVVKADNAA 85



```
RESULT 4
AAY71041
ID AAY71041 standard; peptide; 159 AA.
XX
XX
AC AAY71041;
XX
XX 29-AUG-2000 (first entry)
DT
XX
XX Streptococcus pyogenes strain SF370 GRAB protein fragment #5.
DE
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX
OS Streptococcus pyogenes.
XX
PN WO200026240-A2.
XX
PD 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
PF
XX 02-NOV-1998; 98GB-0023975.
PR
XX (ACTI-) ACTINOVA LTD.
PA
PI Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
DR
XX N-PSDB; AAD00564.
DR
XX
XX New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein -
XX
XX Claim 5; Page 57-58; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding
CC to residues 34-192 and devoid of the membrane spanning region. This
CC fragment is useful in vaccine composition.
XX
XX Query Match 100.0%; Score 129; DB 21; Length 159;
XX Best Local Similarity 100.0%; Pred. No. 8.7e-11;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
DB 59 SDALEALADQTDALQSEEAAYVKADNAA 86
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
RESULT 5
AAY71046
ID AAY71046 standard; Protein; 167 AA.
XX
XX
AC AAY71046;
XX
XX 29-AUG-2000 (first entry)
DT
XX
XX Streptococcus pyogenes strain KTL3 partial GRAB protein.
DE
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX
```

```
OS Streptococcus pyogenes.
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
PF
XX 02-NOV-1998; 98GB-0023975.
PR
XX (ACTI-) ACTINOVA LTD.
PA
PI Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
DR
XX N-PSDB; AAD00564.
DR
XX
XX New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein -
XX
XX Claim 5; Page 62-63; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a partial GRAB protein from S. pyogenes strain KTL3.
CC The protein has alpha2M binding region and is useful in vaccine
CC composition.
XX
XX Query Match 100.0%; Score 129; DB 21; Length 167;
XX Best Local Similarity 100.0%; Pred. No. 9.2e-11;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
DB 70 SDALEALADQTDALQSEEAAYVKADNAA 97
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
RESULT 6
AAY71039
ID AAY71039 standard; Protein; 184 AA.
XX
XX
AC AAY71039;
XX
XX 29-AUG-2000 (first entry)
DT
XX
XX Streptococcus pyogenes strain SF370 mature GRAB protein.
DE
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX
XX Streptococcus pyogenes.
OS
XX
XX Key Location/Qualifiers
FH Binding-site 1..58
FT /label= alpha2-macroglobulin_binding_site
FT Region 59..86
FT /label= R1
FT /note= "repeat region"
FT Region 87..114
FT /label= R2
FT /note= "repeat region"
FT Region 115..159
FT /label= Cell_wall_spanning_region
FT Region 150..155
```

```
FT FT /note= "consensus sequence for gram-positive
FT FT surface cell wall anchored proteins"
FT FT 160..184
FT FT /label= Membrane_spanning_region
XX
PN WO200026240-A2.
XX
XX 11-MAY-2000.
XX
PP 02-NOV-1999; 99WO-GB03631.
XX
PR 02-NOV-1998; 98GB-0023975.
XX
PA (ACTI-) ACTINOVA LTD.
XX
PI Bjorck LH, Rasmussen M;
XX
DR WPI; 2000-365572/31.
XX
PT New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein
XX
PS Claim 5; Page 56; 67pp; English.
XX
CC The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a mature GRAB protein from S. pyogenes strain SF370 without the
CC signal sequence. This sequence is capable of binding alpha2M
XX
XX and useful in vaccine composition.
XX
SQ Sequence 184 AA;
Query Match 100.0%; Score 129; DB 21; Length 184;
Best Local Similarity 100.0%; Pred. No. 1e-10;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SDALEALADQTDALQSEEAAYVKADNAA 28
Db 59 SDALEALADQTDALQSEEAAYVKADNAA 86
RESULT 7
AAY71042
ID AAY71042 standard; Protein: 217 AA.
XX
XX AC AAY71042;
XX
XX 29-AUG-2000 (first entry)
XX
XX Streptococcus pyogenes strain SF370 full-length GRAB protein.
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX
XX Streptococcus pyogenes.
XX
XX OS Location/Qualifiers
XX
XX Key 1..33
XX Peptide /label= Signal_sequence
FT FT 34..217
FT FT /label= Mature_GRAB_protein
FT FT 34..91
FT FT /label= alpha2M_binding_site
FT FT 34..68
FT FT /note= "shows homology to E domain of protein G"
FT FT 92..119
```

```
FT Region /label= Repeat_region_1
FT 120..147
FT /label= Repeat_region_2
FT 148..192
FT /label= Cell_wall_spanning_region
FT 183..188
FT /note= "consensus sequence for gram-positive
FT surface cell wall anchored proteins"
FT 193..217
FT /label= Membrane_spanning_region
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACTI-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
XX
XX N-PSDB; AAD00559; AAD00560.
XX
XX New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein
XX
XX Claim 5; Fig 2B; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G Streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A Streptococcus. Antibodies against GRAB
XX are useful for treating Streptococcus pyogenes infections. The protein
XX is also useful for purifying alpha2M from a sample. The present sequence
XX is a full-length GRAB protein from S. pyogenes strain SF370.
XX
XX SQ Sequence 217 AA;
Query Match 100.0%; Score 129; DB 21; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SDALEALADQTDALQSEEAAYVKADNAA 28
Db 92 SDALEALADQTDALQSEEAAYVKADNAA 119
RESULT 8
AAY71043
ID AAY71043 standard; Protein: 259 AA.
XX
XX AC AAY71043;
XX
XX 29-AUG-2000 (first entry)
XX
XX Streptococcus pyogenes strain KTL9 partial GRAB protein.
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX
XX Streptococcus pyogenes.
XX
XX OS WO200026240-A2.
XX
XX PN 11-MAY-2000.
XX
XX PD 11-MAY-2000.
XX
XX PF 02-NOV-1999; 99WO-GB03631.
```

XX 02-NOV-1998; 98GB-0023975.  
XX (ACTI-) ACTINOVA LTD.  
XX PI Bjorck LH, Rasmussen M;  
XX DR WPI; 2000-365572/31.  
XX DR N-PSDB; AAD00561.  
XX PT New alpha2M binding protein for generating a protective immune response  
XX to group A streptococcus and purifying the binding protein -  
XX  
XX Claim 5; Page 59-60; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
XX related alpha2M binding protein) from Streptococcus pyogenes which have  
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
XX protein G of group G Streptococcus. GRAB protein and peptides derived  
XX from it are used in vaccine compositions for generating a protective  
XX immune response against group A Streptococcus. Antibodies against GRAB  
XX are useful for treating Streptococcus pyogenes infections. The protein  
XX is also useful for purifying alpha2M from a sample. The present sequence  
XX is a partial GRAB protein from S. pyogenes strain KTL9.  
XX The protein has alpha2M binding region and is useful in vaccine  
XX composition.  
XX  
XX SQ . Sequence 259 AA;  
  
Query Match 100.0%; Score 129; DB 21; Length 259;  
Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SDALEALADQTDALQSEAAVVRADNAA 28  
Db 107 SDALEALADQTDALQSEAAVVRADNAA 134  
  
RESULT 9  
AAV71045  
ID AAY71045 standard; Protein; 271 AA.  
XX  
XX AC AAY71045;  
XX  
XX DT 29-AUG-2000 (first entry)  
XX  
XX DE Streptococcus pyogenes strain AP49 partial GRAB protein.  
XX  
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
XX KW immune response; Streptococcus pyogenes infection.  
XX  
XX OS Streptococcus pyogenes.  
XX  
XX PN WO200026240-A2.  
XX  
XX PD 11-MAY-2000.  
XX  
XX PF 02-NOV-1999; 99WO-GB03631.  
XX  
XX PR 02-NOV-1998; 98GB-0023975.  
XX  
XX PA (ACTI-) ACTINOVA LTD.  
XX  
XX PI Bjorck LH, Rasmussen M;  
XX DR WPI; 2000-365572/31.  
XX DR N-PSDB; AAD00563.  
XX  
XX PT New alpha2M binding protein for generating a protective immune response  
XX to group A streptococcus and purifying the binding protein -  
XX  
XX Claim 5; Page 61-62; 67pp; English.

XX The patent discloses a new family of proteins termed GRAB (protein G  
XX related alpha2M binding protein) from Streptococcus pyogenes which have  
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
XX protein G of group G Streptococcus. GRAB protein and peptides derived  
XX from it are used in vaccine compositions for generating a protective  
XX immune response against group A Streptococcus. Antibodies against GRAB  
XX are useful for treating Streptococcus pyogenes infections. The protein  
XX is also useful for purifying alpha2M from a sample. The present sequence  
XX is a partial GRAB protein from S. pyogenes strain AP49.  
XX The protein has alpha2M binding region and is useful in vaccine  
XX composition.  
XX  
XX SQ . Sequence 271 AA;  
  
Query Match 100.0%; Score 129; DB 21; Length 271;  
Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SDALEALADQTDALQSEAAVVRADNAA 28  
Db 174 SDALEALADQTDALQSEAAVVRADNAA 201  
  
RESULT 10  
AAW20445  
ID AAW20445 standard; Protein; 192 AA.  
XX  
XX AC AAW20445;  
XX  
XX DT 11-JUL-1997 (first entry)  
XX  
XX DE H. pylori secreted or periplasmic protein, 35336707.aa.  
XX  
XX KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
XX identification; binding compound; bacterium; life cycle; activator;  
XX KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
XX KW diagnosis.  
XX  
XX OS Helicobacter pylori.  
XX  
XX FH Key Location/Qualifiers  
XX FT Misc-difference 181 /note= "encoded by GYG"  
XX FT Misc-difference 183 /note= "encoded by GYC"  
XX FT Misc-difference 187 /note= "encoded by AWT"  
XX FT Misc-difference 192 /note= "encoded by AWT"  
XX  
XX PN WO9640893-A1.  
XX  
XX PD 19-DEC-1996.  
XX  
XX PF 06-JUN-1996; 96WO-US09122.  
XX  
XX PR 01-APR-1996; 96US-0630405.  
XX PR 07-JUN-1995; 95US-0487032.  
XX  
XX PA (ASTR ) ASTRA AB.  
XX  
XX PI Berglindh OT, Smith D, Mellgaard BL;  
XX  
XX DR WPI; 1997-052306/05.  
XX DR N-PSDB; AAT67618.  
XX  
XX PT Helicobacter pylori nucleic acid sequences and related  
XX polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
XX infection, and to detect Helicobacter  
XX  
XX PS Claim 72; Page 619; 1481pp; English.

This sequence represents a *H. pylori* secreted or periplasmic protein. The protein may be used in a vaccine to prevent or treat *H. pylori* infection or to identify *H. pylori* polypeptide binding compounds, useful as potential *H. pylori* life cycle activators or inhibitors. The genomic sequence of *H. pylori* (ATCC 55679) was determined from overlapping contigs generated by mechanically shearing the bacterial DNA. The sequences were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely *H. pylori* antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified and determined the sequences of interest, particular regions can be isolated from *H. pylori* by PCR amplification for recombinant polypeptide production, e.g. in *E. coli* hosts.

Query Match 44.2%; Score 57; DB 18; Length 192;  
Best Local Similarity 52.2%; Pred. No. 1.8;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY	4	LEALADQTDALQSEEA	VVKADN	26
		: : : :	: : : :	: : : :
Db	49	IOALQEQIDALD	SOEKVSKWDN	71

RESULT 11	
AAW20638	
ID	AAW20638 standard; protein; 222 AA.
XX	
XX	
AC	AAW20638;
XX	
XX	
DT	14-JUL-1997 (first entry)
XX	
DE	H. pylori secreted or periplasmic protein, 02cei02i6orf1.
XX	
XX	
KW	Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW	binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW	duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX	
OS	Helicobacter pylori.

CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide  
CC production, e.g. in *E. coli* hosts.

Query Match 44.2%; Score 57; DB 18; Length 222;  
Best Local Similarity 52.2%; Pred. NO. 2.2;

Qy 4 LEALADQTDLQSEEAUVKADN 26  
::: | | | | : | | |  
pb 77 IOALOEOIDLPSOEKVSKWDN 99

```

RESULT 12
AA10994
ID   AA10994 standard; Protein; 412 AA.
XX
XX   AA10994;
XX
XX   08-JUN-1999   (first entry)
XX
XX
DE   H. pylori ORF 09cel0413_35336707_f2_9 secreted protein..
XX
XX   vaccine; probe; diagnostic; ORF; cell envelope protein;
KW   secreted protein; cellular protein.
KW

```

```
Query Match      44.2%; Score 57; DB 19; Length 412;
Best Local Similarity 52.2%; Pred. No. 4.5;
Matches 12: Conservative 4; Mismatches 7; Indels 0; Gaps 0;
```

[illegible]

OY 4 LEALADQTDALQSEEAHVVKADN 26  
:|||:|||||:|:|  
Db 49 IQALQEOIDALDSQEKVSKWDN 71

RESULT 13  
ABG30268  
ID ABG30268 standard; Protein: 1164 AA.

XX AC ABG30268;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #30259.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Dmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB: AAS94455.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity

XX PS Claim 20; SEQ ID No 60627; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human  
XX CC diagnostic amino acid sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1164 AA;

Query Match 38.8%; Score 50; DB 22; Length 1164;

Best Local Similarity 45.8%; Pred. No. 1.5e+02;

Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 3 ALEALADQTDALQSEEAHVVKADN 26  
:|:|:|:|:|:|:|

Db 581 AVIQLTPQTEAVRSEDAPVAPRDN 604

RESULT 14

AAW11567

ID AAW11567 standard; Protein: 450 AA.

XX AC AAW11567;

XX DT 20-OCT-1997 (first entry)

XX DE E.coli alkaline phosphatase mutant Q329A.

XX KW Bacterial alkaline phosphatase; BAP; mutein; variant; inactive;  
XX KW immunoassay.

XX OS Escherichia coli.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Misc-difference 330

FT /note= "Wild-type Gln329 is replaced by Ala"

XX PN EP752475-A1.

XX PD 08-JAN-1997.

XX PF 27-JUN-1996; 96BP-0401419.

XX PR 29-JUN-1995; 95PR-0007833.

XX PA (BOUL/) BOULAIN J.

XX PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

XX PI Boulain J, Cattolico L, Ducancel F, Menez A;

XX WPI: 1997-067452/07.

XX PT Modified bacterial alkaline phosphatase enzymes - useful as  
XX PT immunoassay labels

XX PS Claim 6; Page -: 35pp; French.

XX CC This sequence represents a specifically claimed modified bacterial  
XX CC alkaline phosphatase (BAP) that has higher activity and similar  
XX CC thermal stability compared with the corresponding wild-type E.coli  
XX CC BAP. In this mutein, one wild-type amino acid residue has been  
XX CC substituted (see features table). Optionally, at least one amino acid  
XX CC is inserted between residues 6 (Pro) and 7 (Val) of the wild-type  
XX CC enzyme. The modified enzyme is useful as an immunoassay label.  
XX CC Note: This sequence does not appear in the specification but has  
XX CC been derived by modifying an E.coli BAP sequence according to the  
XX CC numbering given in the sequence listing.

XX SQ Sequence 450 AA;

Query Match 37.6%; Score 48.5; DB 18; Length 450;

Best Local Similarity 34.1%; Pred. No. 80;

Matches 14; Conservative 6; Mismatches 8; Indels 13; Gaps 1;

OY 1 SDALCALADQTD-----LQSEEAHVVKADNAA 28  
:|:|:|:|:|:|:|

Db 294 NDSVPTLAQMTDAIKELLSKNEKGFQFVQEGASIDKADHAA 334  
:|:|:|:|:|:|:|

RESULT 15

AAV36771

ID AAV36771 standard; Protein: 258 AA.

XX AC AAV36771;

XX DT 07-OCT-1999 (first entry)



XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-483447/52.  
DR Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -  
XX  
XX Claim 27; SEQ ID NO 35991; 639pp + sequence listing; English.  
XX The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 73 AA;  
XX  
XX Query Match 35.3%; Score 45.5; DB 22; Length 73;  
XX Best Local Similarity 48.1%; Pred. No. 26;  
XX Matches 13; Conservative 2; Mismatches 7; Indels 5; Gaps 1;  
XX  
XX QY 2 DALEALADQTDALQS-----EEAAVVK 23  
XX | | | | | | | | | | : | : | |  
XX Db 47 DILEQTAQAQVDALQGHLLQMQEALVK 73  
XX  
XX RESULT 18  
XX AAM64267  
XX ID AAM64267 standard; Protein; 73 AA.  
XX AC AAM64267;  
XX XX  
XX DT 05-NOV-2001 (first entry)  
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 36372.  
XX KW Human; brain expressed exon; gene expression analysis; probe;  
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX KW epilepsy; cancer.  
XX OS Homo sapiens.  
XX PN WO200157275-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00667.  
XX XX  
XX 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PI WPI; 2001-483446/52.  
XX DR  
XX XX  
XX 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PI WPI; 2001-483446/52.  
XX DR  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX PT  
XX PS Example 4; SEQ ID NO: 36372; 650pp + Sequence Listing; English.  
XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, of  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.  
XX  
XX Sequence 73 AA;  
XX  
XX Query Match 35.3%; Score 45.5; DB 22; Length 73;  
XX Best Local Similarity 48.1%; Pred. No. 26;  
XX Matches 13; Conservative 2; Mismatches 7; Indels 5; Gaps 1;  
XX  
XX QY 2 DALEALADQTDALQS-----EEAAVVK 23  
XX | | | | | | | | | | : | : | |  
XX Db 47 DILEQTAQAQVDALQGHLLQMQEALVK 73  
XX  
XX RESULT 19  
XX AAM77089  
XX ID AAM77089 standard; Protein; 73 AA.  
XX AC AAM77089;  
XX XX  
XX DT 06-NOV-2001 (first entry)  
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 37395.  
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;  
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX OS Homo sapiens.  
XX PN WO200157276-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00668.  
XX XX  
XX 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PI WPI; 2001-488900/53.  
XX DR  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow -  
XX PS  
XX Example 4; SEQ ID NO: 37395; 658pp + Sequence Listing; English.  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukaemia and myeloma. The present sequence is a  
XX protein encoded by one of the probes of the invention.  
XX  
XX Sequence 73 AA;  
XX  
XX Query Match 35.3%; Score 45.5; DB 22; Length 73;  
XX Best Local Similarity 48.1%; Pred. No. 26;  
XX Matches 13; Conservative 2; Mismatches 7; Indels 5; Gaps 1;  
XX  
XX QY 2 DALEALADQTDALQS-----EEAAVVK 23  
XX | | | | | | | | | | : | : | |









PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 14-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 35.3%; Score 45.5; DB 21; Length 572;  
 Best Local Similarity 61.9%; Pred. No. 2.8e+02;  
 Matches 13; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 4 LEALADQTDALQSEEAHVRA 24  
 ||||:| | | | | | | |  
 Db 91 LEALASET-LLTSDEAVVAA 110

RESULT 23  
 AAU34319  
 ID AAU34319 standard; Protein; 2076 AA.

XX AC AAU34319;  
 XX DT 14-FEB-2002 (first entry)  
 XX DE Staphylococcus aureus cellular proliferation protein #595.  
 XX KW Antisense; prokaryotic cellular proliferation protein;  
 XX KW antibiotic; antibacterial; drug design.

XX .OS Staphylococcus aureus.  
 XX PN WO200170955-A2.  
 XX PD 27-SEP-2001.

XX XX 21-MAR-2001; 2001WO-US09180.  
 XX XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WI WPI; 2001-611495/70.  
 DR N-PSDB; AAS521178.

XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX Example 3; Seq ID No 5815; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 2076 AA;

Query Match 35.3%; Score 45.5; DB 22; Length 2076;  
 Best Local Similarity 48.0%; Pred. No. 1.3e+03;  
 Matches 12; Conservative 3; Mismatches 9; Indels 1; Gaps 1;

QY 2 DALEALA-DQTDALQSEEAHVRA 25  
 ||||| | | | | | | | |  
 Db 766 DALNQLATDETDIDVNTNATNAD 790

RESULT 24  
 AAU37320  
 ID AAU37320 standard; Protein; 2186 AA.

XX AC AAU37320;  
 XX DT 14-FEB-2002 (first entry)  
 XX DE Staphylococcus aureus cellular proliferation protein #1490.  
 XX KW Antisense; prokaryotic cellular proliferation protein;  
 XX KW antibiotic; antibacterial; drug design.

XX OS Staphylococcus aureus.  
 XX PN WO200170955-A2.  
 XX PD 27-SEP-2001.

XX XX 21-MAR-2001; 2001WO-US09180.  
 XX XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.



FT Domain 227..234 /note= "Ser/Thr kinase motif"  
FT Domain 346..364 /note= "regulatory domain"  
FT  
XX EP978562-A1.  
PN  
XX 09-FEB-2000.  
XX  
XX 06-AUG-1998; 98EP-0202654.  
XX  
XX 06-AUG-1998; 98EP-0202654.  
XX  
PA (MEDI-) DIV MEDICAL PHARMACOLOGY LEIDEN AMSTERDA.  
XX  
XX WPI; 2000-138770/13.  
XX  
XX Mammalian calcium/calmodulin dependent protein kinase related protein  
PT useful as a modulator of calcium/calmodulin dependent protein kinase  
PT activity, useful for treating seizures and damage to the central  
PT nervous system -  
XX  
PS Disclosure; Page 34-36; 48pp; English.  
XX  
CC The present sequence represents a calcium/calmodulin dependent protein  
CC kinase (CaMK-IV). The specification also describes another CaMK,  
CC designated CaMK-VI, and a CaMK-VI related peptide, designated CARP. The  
CC CARP peptide is produced by alternative splicing of the CaMK-VI gene,  
CC and plays a role in kinase-mediated neuronal plasticity which is  
CC associated with epilepsy. A large part of CARP (48 out of 55 amino  
CC acids) is highly homologous with the carboxy-terminus of human  
CC doublecortin. It is possible the calcium overload in a cell, triggered  
CC by kainate-induced seizures, leads to activation of CaMK-VI. This  
CC enhanced kinase activity induces CARP expression, which subsequently  
CC acts as a modulator of the kinase, e.g. by acting as a substrate or  
CC by irreversible binding to the activated kinase. The mammalian CARP  
CC protein may be used for the treatment and/or prevention of seizures  
CC (e.g. in epilepsy) and/or damage to the central nervous system,  
CC particularly the hippocampus. CARP peptides may also be used to treat  
CC acute brain insults (e.g. resulting from stroke, trauma, hypoxia, and  
CC ischemia), and neurodegenerative diseases (e.g. Alzheimer's disease).  
XX  
SQ Sequence 502 AA;  
Query Match 34.9%; Score 45; DB 21; Length 502;  
Best Local Similarity 58.8%; Pred. No. 2.9e+02;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 2 DALEALADQTDALQSEE 18  
: : : : :  
Db 406 EAAKAAADETMKLQSEE 422  
RESULT 27  
AAW99481  
ID AAW99481 standard; Protein; 1362 AA.  
XX  
AC AAW99481;  
XX  
XX 08-JUN-1999 (first entry)  
XX  
XX Murine pCIP protein.  
XX  
XX Mouse; pCIP; p300/CBP/co-Integrator-associated protein; gene expression;  
KW breast cancer; inflammatory disease; atherosclerosis; osteoporosis; ds.  
XX  
XX Mus sp.  
XX  
XX Key Location/Qualifiers  
PH  
FT Misc-difference 540 /note= "defined in specification as ?"  
FT Misc-difference 1080 /note= "defined in specification as O"  
FT

FT Misc-difference 44..45 /note= "there appears to be 20 amino acids missing  
FT between these positions in the sequence given  
FT in the specification"  
FT  
FT Misc-difference 1044..1045 /note= "there appears to be 20 amino acids missing  
FT between these positions in the sequence given  
FT in the specification"  
XX  
PN W09856806-A1.  
XX  
XX 17-DEC-1998.  
XX  
XX 12-JUN-1998; 98WO-US12263.  
XX  
XX 12-JUN-1997; 97US-0049452.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Glass CK, Rose DW, Rosenfield MG, Torchia J;  
PI  
XX WPI; 1999-080883/07.  
DR N-PSDB; AAX26000.  
XX  
XX New nucleic acids encoding pCIP and NcoA-2 polypeptides - are used  
PT to identify agents that regulate gene expression, e.g. for treatment  
PT of cancer, inflammatory disease and osteoporosis  
XX  
PS Claim 16; Fig 1; 100pp; English.  
XX  
CC This sequence represents the amino acid sequence of the mouse pCIP  
CC (p300/CBP/co-integrator-associated protein) polypeptide. The protein  
CC can regulate gene expression so are potentially useful therapeutically,  
CC e.g. against (breast) cancer, inflammatory disease (e.g. atherosclerosis)  
CC or osteoporosis.  
XX  
SQ Sequence 1362 AA;  
Query Match 34.9%; Score 45; DB 20; Length 1362;  
Best Local Similarity 41.7%; Pred. No. 9.1e+02;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;  
QY 5 EALADQTDALQSEEAAYVKADNAA 28  
: : : : :  
Db 1053 QALESKODVFQGEAAVMDQKAA 1076  
RESULT 28  
AAW81029  
ID AAW81029 standard; Protein; 1447 AA.  
XX  
XX AAW81029;  
AC  
XX  
XX 12-MAY-1999 (first entry)  
XX  
XX Murine pCIP protein (ortholog of human AIB1 protein).  
XX  
XX AIB1; amplified in breast cancer; cancer; steroid; receptor;  
KW coactivator; SCR; estrogen; ER; estrogen dependent transcription;  
KW breast cancer; lung cancer; colon cancer; prostate cancer;  
KW melanoma.  
XX  
XX Mus musculus.  
XX  
XX Key Location/Qualifiers  
FH Region 1402  
FT /label= Potential\_C\_terminus  
FT /note= "this position encoded by stop codon"  
XX  
XX W09857982-A2.  
XX  
XX 23-DEC-1998.  
XX



CC nucleotide sequence selected from AAV43552 to AAV43560 and nucleotide  
 CC sequences identical to a known splice branch site in a known gene,  
 CC sequences identical in length to a known spacer region between splice  
 CC branch and acceptor sites in a known gene, sequences identical to a known  
 CC splice acceptor site in a known gene, sequence identical to a known  
 CC splice donor site in a known gene, an open reading frame (ORF) 3N-1  
 CC nucleotides in length, the ORF encoding a known peptide tag recognisable  
 CC by a known reaction characteristic of the known peptide tag and sequences  
 CC selected from CAGS and TAGG. The DNA sequence is inserted into the intron  
 CC within the gene to create a tagged gene, and the tagged gene is incubated  
 CC within a cell so as to maintain intact or to introduce the tagged gene  
 CC within the genome of the cell. The method is used for isolating proteins,  
 CC RNA and genes, for analysis of subcellular structures, cellular responses  
 CC and transcriptional regulation, for the study of viral infection and for  
 CC diagnosis of disease.

XX Sequence 515 AA;

Query Match 34.5%; Score 44.5; DB 19; Length 515;  
 Best Local Similarity 51.7%; Pred. No. 3.5e+02;  
 Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

QY 1 SDALEALADQTDALQSEEAAYVK--ADNA 27  
 :||:||||| : | :||| ||| | ||| |  
 Db 368 ADAVEAAAEELTA-QAEAAAKWEADKA 395

RESULT 31  
 AAW70356

ID AAW70356 standard; Protein; 516 AA.

XX AC AAW70356;

XX DT 14-DEC-1998 (first entry)

XX DE Chlamydomonas protein RSP3.

XX KW Target nucleotide; epitope; gene tagging; epitope tagging.

XX OS Chlamydomonas sp.

XX PN WO9826094-A1.

XX PD 18-JUN-1998.

XX PF 09-DEC-1997; 97WO-US22472.

XX PR 09-DEC-1996; 96US-0762106.

XX PA (JARV/) JARVIK J W.

XX PI Jarvik JW;

XX DR WPI; 1998-348546/30.

XX DR N-PSDB; AAV33277.

XX PT Epitope tagging of genes, transcripts and polypeptides - using an  
 PT oligonucleotide comprising a nucleotide sequence encoding the  
 PT epitope independently of the reading frame of the nucleotide  
 PT sequence

XX PS Disclosure; Fig 2A; 46pp; English.

XX CC The Chlamydomonas protein RSP3 is an example of a target nucleotide  
 CC sequence into which is inserted an oligonucleotide encoding an epitope.  
 CC The epitope is encoded by the oligonucleotides nucleotide sequence  
 CC independently of the reading frame. The epitope and methods disclosed  
 CC can be used for the tagging of genes, transcripts and polypeptides. They  
 CC can be used for e.g. discovering new genes, determining the size and  
 CC abundance of proteins produced by newly discovered genes, tracking the  
 CC movement of proteins within cell membranes, monitoring receptor binding  
 CC and internalisation of exogenous proteins, identifying the components of  
 CC functional protein complexes, purifying proteins, discovering the

CC function of proteins. The products, methods and uses are particularly  
 CC used for proteins that are unstable, are difficult to purify, or share  
 CC epitopes with a number of other proteins. The methods overcome the  
 CC inefficiency of epitope tagging caused by reading frame obstacles and  
 CC orientation obstacles.

XX SQ Sequence 516 AA;

Query Match 34.5%; Score 44.5; DB 19; Length 516;  
 Best Local Similarity 51.7%; Pred. No. 3.5e+02;  
 Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

QY 1 SDALEALADQTDALQSEEAAYVK--ADNA 27  
 :||:||||| : | :||| ||| | ||| |  
 Db 369 ADAVEAAAEELTA-QAEAAAKWEADKA 396

RESULT 32

AAW70357

ID AAW70357 standard; Protein; 527 AA.

XX AC AAW70357;

XX DT 14-DEC-1998 (first entry)

XX DE Tagged Chlamydomonas protein RSP3.

XX KW Target nucleotide; epitope; gene tagging; epitope tagging.

XX OS Chlamydomonas sp.

XX FH Key Location/Qualifiers

XX FT Region 147..157

XX FT /note="Tag"

XX PN WO9826094-A1.

XX PD 18-JUN-1998.

XX PF 09-DEC-1997; 97WO-US22472.

XX PR 09-DEC-1996; 96US-0762106.

XX PA (JARV/) JARVIK J W.

XX PI Jarvik JW;

XX DR WPI; 1998-348546/30.

XX DR N-PSDB; AAV33278.

XX PT Epitope tagging of genes, transcripts and polypeptides - using an  
 PT oligonucleotide comprising a nucleotide sequence encoding the  
 PT epitope independently of the reading frame of the nucleotide  
 PT sequence

XX PS Disclosure; Fig 2B; 46pp; English.

XX CC The tagged Chlamydomonas protein RSP3 is an example of a target  
 CC nucleotide with an oligonucleotide encoding an epitope inserted. The  
 CC epitope is encoded by the oligonucleotides nucleotide sequence  
 CC independently of the reading frame. The epitope and methods disclosed  
 CC can be used for the tagging of genes, transcripts and polypeptides. They  
 CC can be used for e.g. discovering new genes, determining the size and  
 CC abundance of proteins produced by newly discovered genes, tracking the  
 CC movement of proteins within cell membranes, monitoring receptor binding  
 CC and internalisation of exogenous proteins, identifying the components of  
 CC functional protein complexes, purifying proteins, discovering the  
 CC function of proteins. The products, methods and uses are particularly  
 CC used for proteins that are unstable, are difficult to purify, or share  
 CC epitopes with a number of other proteins. The methods overcome the  
 CC inefficiency of epitope tagging caused by reading frame obstacles and  
 CC orientation obstacles.

SQ Sequence 527 AA;

Query Match 34.5%; Score 44.5; DB 19; Length 527;  
Best Local Similarity 51.7%; Pred. No. 3.6e+02;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Oy 1 SDALFALADQTDALQSEEAAYVK--ADNA 27  
:||:| | : | :||| | | |  
Db 380 ADAVEAAALTA-QAEAAANKWEADKA 407

RESULT 33  
AAW61908  
ID AAW61908 standard; Protein; 552 AA.  
XX  
AC AAW61908;  
XX  
DT 16-SEP-1998 (first entry)  
XX  
DE Tandem tagged RSP3 pf14 gene product; protein.  
XX  
KW Tagged gene; tagged transcript; hybrid intron; protein tag;  
KW protein isolation; recombination; subcellular structure analysis;  
KW transcriptional regulation; viral infection; Chlamydomonas; pf14 gene.  
XX  
OS Chlamydomonas sp.  
OS Synthetic.  
XX  
PN WO9820031-A1.  
XX  
PD 14-MAY-1998.  
XX  
PF 07-NOV-1997; 97WO-US20150.  
XX  
PR 08-NOV-1996; 96US-0705404.  
XX  
PA (JARV/) JARVIK J W.  
XX  
PI Jarvik JW;  
XX  
DR WPT; 1998-286861/25.  
XX  
PT Tagging genes, transcripts and proteins - using tag-creating DNA  
PT inserted into intron of gene to create 2 hybrid introns separated by  
PT new exon encoding protein tag  
XX  
PS Disclosure; Pages 22-23; 66pp; English.  
XX  
CC This sequence represents a tandem tagged RSP3 protein, a Chlamydomonas  
CC pf14 gene product. The gene can be tagged by using the method of  
CC invention of tagging genes, transcripts and proteins in cells in a  
CC single recombinational event. The method comprises producing a tagged  
CC gene by inserting a DNA sequence into an intron of a gene by selecting a  
CC DNA sequence having a 5' portion free of any nucleotide sequence  
CC selected from AAV43548 to AAV43551, a nucleotide sequence selected from  
CC AAV43552 to AAV43560 and nucleotide sequences identical to a known splice  
CC branch site in a known gene, sequences identical in length to a known  
CC spacer region between splice branch and acceptor sites in a known gene,  
CC sequences identical to a known splice donor site in a known gene,  
CC sequence identical to a known splice donor site in a known gene, an open  
CC reading frame (ORF) 3N-1 nucleotides in length, the ORF encoding a known  
CC peptide tag recognizable by a known reaction characteristic of the known  
CC peptide tag and sequences selected from CAGG and TAGG. The DNA sequence  
CC is inserted into the intron within the gene to create a tagged gene, and  
CC the tagged gene is incubated within a cell so as to maintain intact or  
CC to introduce the tagged gene within the genome of the cell. The method  
CC is used for isolating proteins, RNA and genes, for analysis of  
CC subcellular structures, cellular responses and transcriptional  
CC regulation, for the study of viral infection and for diagnosis of  
CC disease.  
XX  
SQ Sequence 552 AA;

Query Match 34.5%; Score 44.5; DB 19; Length 552;  
Best Local Similarity 51.7%; Pred. No. 3.8e+02;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Oy 1 SDALFALADQTDALQSEEAAYVK--ADNA 27  
:||:| | : | :||| | | |  
Db 405 ADAVEAAALTA-QAEAAANKWEADKA 432

RESULT 34  
ABG01077  
ID ABG01077 standard; Protein; 42 AA.  
XX  
AC ABG01077;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #1068.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
XX  
DR N-PSDB; AAS65264.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID NO 31436; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG03077 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_sequences.  
XX  
SQ Sequence 42 AA;

Query Match 34.1%; Score 44; DB 22; Length 42;  
Best Local Similarity 36.4%; Pred. No. 22;



	Matches	8;	Conservative	7;	Mismatches	7;	Indels	0;	Gaps	0;
QY	2	DALEALADOTDALQSEAAVVK	23							
		l:l: l:l: l:l: l:l: l:l:								
Db	8	DSLQLLODEDSANEEGLVLR	29							
RESULT 35										
AAAG58293	ID AAG58293 standard; Protein; 274 AA.									
XX	AC	AAG58293;								
XX	DT	18-OCT-2000 (first entry)								
DE	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 75231.								
XX	KW	Protein identification; signal transduction pathway; metabolic pathway;								
KW	KW	hybridisation assay; genetic mapping; gene expression control; promoter;								
KW		termination sequence.								
OS	XX	Arabidopsis thaliana.								
XX	PN	EP1033405-A2.								
XX	PD	06-SEP-2000.								
XX	PF	25-FEB-2000; 2000EP-0301439.								
XX	PR	25-FEB-1999; 99US-0121825.								
PR	PR	05-MAR-1999; 99US-0123180.								
PR	PR	09-MAR-1999; 99US-0123548.								
PR	PR	23-MAR-1999; 99US-0125788.								
PR	PR	25-MAR-1999; 99US-0126264.								
PR	PR	29-MAR-1999; 99US-0126785.								
PR	PR	01-APR-1999; 99US-0127462.								
PR	PR	06-APR-1999; 99US-0128234.								
PR	PR	16-APR-1999; 99US-0128714.								
PR	PR	19-APR-1999; 99US-0129845.								
PR	PR	21-APR-1999; 99US-0130077.								
PR	PR	23-APR-1999; 99US-0130449.								
PR	PR	23-APR-1999; 99US-0130510.								
PR	PR	28-APR-1999; 99US-0130891.								
PR	PR	30-APR-1999; 99US-0131449.								
PR	PR	04-MAY-1999; 99US-0132048.								
PR	PR	04-MAY-1999; 99US-0132484.								
PR	PR	05-MAY-1999; 99US-0132485.								
PR	PR	06-MAY-1999; 99US-0132486.								
PR	PR	07-MAY-1999; 99US-0132487.								
PR	PR	11-MAY-1999; 99US-0132863.								
PR	PR	14-MAY-1999; 99US-0134218.								
PR	PR	14-MAY-1999; 99US-0134219.								
PR	PR	14-MAY-1999; 99US-0134221.								
PR	PR	18-MAY-1999; 99US-0134370.								
PR	PR	19-MAY-1999; 99US-0134768.								
PR	PR	20-MAY-1999; 99US-0134941.								
PR	PR	21-MAY-1999; 99US-0135124.								
PR	PR	24-MAY-1999; 99US-0135353.								
PR	PR	25-MAY-1999; 99US-0135629.								
PR	PR	27-MAY-1999; 99US-0136021.								
PR	PR	28-MAY-1999; 99US-0136392.								
PR	PR	01-JUN-1999; 99US-0137222.								
PR	PR	03-JUN-1999; 99US-0137528.								
PR	PR	04-JUN-1999; 99US-0137502.								
PR	PR	07-JUN-1999; 99US-0137724.								
PR	PR	08-JUN-1999; 99US-0138094.								
PR	PR	10-JUN-1999; 99US-0138540.								
PR	PR	10-JUN-1999; 99US-0138847.								
PR	PR	14-JUN-1999; 99US-0139119.								
PR	PR	16-JUN-1999; 99US-0139452.								
PR	PR	16-JUN-1999; 99US-0139453.								
PR	PR	17-JUN-1999; 99US-0139492.								
PR	PR	18-JUN-1999; 99US-0139454.								
PR	PR	18-JUN-1999; 99US-0139455.								
PR	PR	18-JUN-1999; 99US-0139456.								
PR	PR	18-JUN-1999; 99US-0139457.								
PR	PR	18-JUN-1999; 99US-0139458.								
PR	PR	18-JUN-1999; 99US-0139459.								
PR	PR	18-JUN-1999; 99US-0139460.								
PR	PR	18-JUN-1999; 99US-0139461.								
PR	PR	18-JUN-1999; 99US-0139462.								
PR	PR	18-JUN-1999; 99US-0139463.								
PR	PR	18-JUN-1999; 99US-0139750.								
PR	PR	18-JUN-1999; 99US-0139763.								
PR	PR	21-JUN-1999; 99US-0139817.								
PR	PR	22-JUN-1999; 99US-0139899.								
PR	PR	23-JUN-1999; 99US-0140353.								
PR	PR	23-JUN-1999; 99US-0140354.								
PR	PR	24-JUN-1999; 99US-0140695.								
PR	PR	28-JUN-1999; 99US-0140823.								
PR	PR	29-JUN-1999; 99US-0140991.								
PR	PR	30-JUN-1999; 99US-0141287.								
PR	PR	01-JUL-1999; 99US-0141842.								
PR	PR	01-JUL-1999; 99US-0142154.								
PR	PR	02-JUL-1999; 99US-0142055.								
PR	PR	06-JUL-1999; 99US-0142390.								
PR	PR	08-JUL-1999; 99US-0142803.								
PR	PR	09-JUL-1999; 99US-0142920.								
PR	PR	12-JUL-1999; 99US-0142977.								
PR	PR	13-JUL-1999; 99US-0143542.								
PR	PR	14-JUL-1999; 99US-0143624.								
PR	PR	15-JUL-1999; 99US-0144005.								
PR	PR	16-JUL-1999; 99US-0144085.								
PR	PR	16-JUL-1999; 99US-0144086.								
PR	PR	19-JUL-1999; 99US-0144325.								
PR	PR	19-JUL-1999; 99US-0144331.								
PR	PR	19-JUL-1999; 99US-0144332.								
PR	PR	19-JUL-1999; 99US-0144333.								
PR	PR	19-JUL-1999; 99US-0144334.								
PR	PR	19-JUL-1999; 99US-0144335.								
PR	PR	20-JUL-1999; 99US-0144352.								
PR	PR	20-JUL-1999; 99US-0144632.								
PR	PR	21-JUL-1999; 99US-0144884.								
PR	PR	21-JUL-1999; 99US-0144814.								
PR	PR	21-JUL-1999; 99US-0145086.								
PR	PR	21-JUL-1999; 99US-0145088.								
PR	PR	22-JUL-1999; 99US-0145085.								
PR	PR	22-JUL-1999; 99US-0145087.								
PR	PR	22-JUL-1999; 99US-0145089.								
PR	PR	22-JUL-1999; 99US-0145192.								
PR	PR	23-JUL-1999; 99US-0145145.								
PR	PR	23-JUL-1999; 99US-0145218.								
PR	PR	23-JUL-1999; 99US-0145224.								
PR	PR	26-JUL-1999; 99US-0145276.								
PR	PR	27-JUL-1999; 99US-0145913.								
PR	PR	27-JUL-1999; 99US-0145918.								
PR	PR	27-JUL-1999; 99US-0145919.								
PR	PR	28-JUL-1999; 99US-0145951.								
PR	PR	02-AUG-1999; 99US-0146386.								
PR	PR	02-AUG-1999; 99US-0146388.								
PR	PR	02-AUG-1999; 99US-0146389.								
PR	PR	03-AUG-1999; 99US-0147038.								
PR	PR	04-AUG-1999; 99US-0147204.								
PR	PR	04-AUG-1999; 99US-0147302.								
PR	PR	05-AUG-1999; 99US-0147192.								
PR	PR	05-AUG-1999; 99US-0147260.								
PR	PR	06-AUG-1999; 99US-0147303.								
PR	PR	06-AUG-1999; 99US-0147416.								
PR	PR	09-AUG-1999; 99US-0147493.								
PR	PR	09-AUG-1999; 99US-0147935.								
PR	PR	10-AUG-1999; 99US-0148171.								
PR	PR	11-AUG-1999; 99US-0148319.								
PR	PR	12-AUG-1999; 99US-0148341.								



PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140595.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 34.1%; Score 44; DB 21; Length 274;  
Best Local Similarity 39.1%; Pred. No. 2e+02;  
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADOTDALQSEAAVVKADNA 27  
||| : || : ||:: |||  
Db 53 EALQQRDKALSERDKALIERDNA 75

RESULT 37  
AAM78512  
ID AAM78512 standard; Protein; 284 AA.  
XX  
AC AAM78512;  
XX  
DT 06-NOV-2001 (first entry)

XX	Human protein SEQ ID NO 1174.	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 75230.
DE		XX	
XX	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;	KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;	XX	termination sequence.
KW	nervous system disorder; arthritis; inflammation.	OS	Arabidopsis thaliana.
XX		XX	
OS	Homo sapiens.	XX	
XX		PN	EP1033405-A2.
PN	WO200157190-A2.	PN	
XX		PD	06-SEP-2000.
PD	09-AUG-2001.	XX	
XX		PF	25-FEB-2000; 2000EP-0301439.
PF	05-FEB-2001; 2001WO-US04098.	XX	
XX		PR	25-FEB-1999; 99US-0121825.
PR	03-FEB-2000; 2000US-0496914.	PR	05-MAR-1999; 99US-0123180.
PR	27-APR-2000; 2000US-0560875.	PR	09-MAR-1999; 99US-0123548.
PR	20-JUN-2000; 2000US-0598075.	PR	23-MAR-1999; 99US-0125788.
PR	19-JUL-2000; 2000US-0620325.	PR	25-MAR-1999; 99US-0126264.
PR	01-SEP-2000; 2000US-0654936.	PR	29-MAR-1999; 99US-0126785.
PR	15-SEP-2000; 2000US-0663561.	PR	01-APR-1999; 99US-0127462.
PR	20-OCT-2000; 2000US-0693325.	PR	06-APR-1999; 99US-0128234.
PR	30-NOV-2000; 2000US-0728422.	PR	08-APR-1999; 99US-0128714.
XX		PR	16-APR-1999; 99US-0129845.
XX	(HYSE-) HYSEQ INC.	PR	19-APR-1999; 99US-0130077.
PA		PR	21-APR-1999; 99US-0130449.
XX	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;	PR	23-APR-1999; 99US-0130510.
PI	Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;	PR	23-APR-1999; 99US-0130891.
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;	PR	28-APR-1999; 99US-0131449.
XX		PR	30-APR-1999; 99US-0132048.
DR	WPI; 2001-476283/51.	PR	04-MAY-1999; 99US-0132484.
DR	N-PSDB; AAK51645.	PR	05-MAY-1999; 99US-0132485.
XX		PR	06-MAY-1999; 99US-0132486.
PT	Nucleic acids encoding polypeptides with cytokine-like activities,	PR	06-MAY-1999; 99US-0132487.
PT	useful in diagnosis and gene therapy -	PR	07-MAY-1999; 99US-0132863.
XX		PR	11-MAY-1999; 99US-0134256.
PS	Claim 20; Page 3417-3418; 6221pp; English.	PR	14-MAY-1999; 99US-0134218.
XX		PR	14-MAY-1999; 99US-0134219.
CC	The invention relates to polynucleotides (AAK51456-AAK53435) and the	PR	14-MAY-1999; 99US-0134221.
CC	encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to	PR	14-MAY-1999; 99US-0134370.
CC	cytokine, cell proliferation or cell differentiation or which may induce	PR	18-MAY-1999; 99US-0134766.
CC	production of other cytokines in other cell populations. The	PR	19-MAY-1999; 99US-0134941.
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or	PR	20-MAY-1999; 99US-0135124.
CC	peptide therapy. The polypeptides have various cytokine-like activities,	PR	21-MAY-1999; 99US-0135353.
CC	e.g. stem cell growth factor activity, haematopoiesis regulating	PR	24-MAY-1999; 99US-0135629.
CC	activity, tissue growth factor activity, immunomodulatory activity and	PR	25-MAY-1999; 99US-0136021.
CC	activin/inhibin activity and may be useful in the diagnosis and/or	PR	27-MAY-1999; 99US-0136392.
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and	PR	27-MAY-1999; 99US-0136782.
CC	inflammation.	PR	01-JUN-1999; 99US-0137222.
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666	PR	03-JUN-1999; 99US-0137528.
CC	(AAM80020) are omitted as the relevant pages from the sequence listing	PR	04-JUN-1999; 99US-0137502.
CC	were missing at the time of publication.	PR	07-JUN-1999; 99US-0137724.
XX		PR	08-JUN-1999; 99US-0138094.
SQ	Sequence 284 AA;	PR	10-JUN-1999; 99US-0138540.
		PR	10-JUN-1999; 99US-0138847.
	Query Match 34.1%; Score 44; DB 22; Length 284;	PR	14-JUN-1999; 99US-0139119.
	Best Local Similarity 47.6%; Pred. No. 2e+02;	PR	16-JUN-1999; 99US-0139452.
	Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;	PR	16-JUN-1999; 99US-0139453.
		PR	17-JUN-1999; 99US-0139492.
QY	5 EALADOTDALQSEEAAYVKAD 25	PR	18-JUN-1999; 99US-0139454.
	: : : : : :	PR	18-JUN-1999; 99US-0139455.
DB	62 EALKDAEDSLAAEEAAAKAE 82	PR	18-JUN-1999; 99US-0139456.
		PR	18-JUN-1999; 99US-0139457.
		PR	18-JUN-1999; 99US-0139458.
		PR	18-JUN-1999; 99US-0139459.
RESULT 38		PR	18-JUN-1999; 99US-0139460.
AGS58292		PR	18-JUN-1999; 99US-0139461.
ID	AGS58292 standard; Protein; 296 AA.	PR	18-JUN-1999; 99US-0139462.
XX		PR	18-JUN-1999; 99US-0139463.
AC	AGS58292;	PR	18-JUN-1999; 99US-0139750.
XX		PR	18-JUN-1999; 99US-0139753.
DT	18-OCT-2000 (first entry)	PR	21-JUN-1999; 99US-0139817.
XX			







THIS PAGE BLANK (uspto)



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:50:13 ; Search time 1877.39 seconds  
(without alignments)  
1143.085 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQRIIPNGTTLNL.....KKEERQNVNLTPTTGESNP 159

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q/cgn2\_1/USPTO\_spool/US09847539/runat\_10102002\_093106\_5005/app\_query.fasta\_1.526  
-DB-EST -OPMT-fastap -SUFFIX-std.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE-LOCAL  
-OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER-US09847539\_6CGN\_1\_1\_763 @runat\_10102002\_093106\_5005 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FPGAPOP=6 -FPGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	127.5	16.4	938	12 CNS07CG2	AL439128 T7 end of
2	112.5	14.5	689	10 BF480699	BF480699 L0-2660T3

c 3	112.5	14.5	1061	12 Cns0738M	AL427196 clone BA0
c 4	112	14.4	442	9 AU180004	AU180004 AU180004
c 5	112	14.4	475	10 BM029212	BM029212 IpSkn0005
c 6	112	14.4	570	9 AU179963	AU179963 AU179963
c 7	111.5	14.4	615	9 AU057981	AU057981 SMOVL3CAN
c 8	111	14.3	344	9 AU180137	AU180137 AU180137
c 9	110.5	14.2	617	10 BM355271	BM355271 rr22h07.y
c 10	110	14.2	517	9 AU179732	AU179732 AU179732
c 11	110	14.2	610	9 AU180091	AU180091 AU180091
c 12	110	14.2	610	9 AU180168	AU180168 AU180168
c 13	109.5	14.1	453	9 AU179269	AU179269 AU179269
c 14	108.5	14.0	715	10 BJ064193	BJ064193 BJ064193
c 15	108	13.9	548	9 AU177381	AU177381 AU177381
c 16	107	13.8	380	9 AU179899	AU179899 AU179899
c 17	107	13.8	664	10 BG443972	BG443972 GA_Ea002
c 18	107	13.8	857	10 BM360707	BM360707 GA_Ea003
c 19	107	13.8	881	10 BG445649	BG445649 GA_Ea002
c 20	107	13.8	887	10 BG444821	BG444821 GA_Ea002
c 21	106	13.6	546	9 AI959039	AI959039 fd06g08.y
c 22	106	13.6	620	10 BJ114733	BJ114733 BJ114733
c 23	106	13.6	947	12 Cns0794Y	AL434840 T3 end of
c 24	105.5	13.6	653	10 BJ001775	BJ001775 BJ001775
c 25	105.5	13.6	669	10 BJ000182	BJ000182 BJ000182
c 26	105.5	13.6	687	10 BJ011593	BJ011593 BJ011593
c 27	105.5	13.6	710	10 BJ001283	BJ001283 BJ001283
c 28	105.5	13.6	968	12 Cns07CXJ	AL439757 T7 end of
c 29	105	13.5	473	10 BE405838	BE405838 WHE0437.A
c 30	105	13.5	755	9 AJ395963	AJ395963 AJ395963
c 31	104	13.4	530	10 BM300555	BM300555 MCA053H02
c 32	103.5	13.3	616	10 BJ006142	BJ006142 BJ006142
c 33	103.5	13.3	666	10 BJ006464	BJ006464 BJ006464
c 34	103	13.3	938	12 A2676392	A2676392 ENTHV47TR
c 35	102.5	13.2	537	10 BG930118	BG930118 etsHES05
c 36	102.5	13.2	628	12 A0648620	A0648620 RPC193-BC
c 37	102.5	13.2	790	12 Cns07270	AL425866 clone BA0
c 38	102	13.1	727	10 BG440614	BG440614 GA_Ea000
c 39	102	13.1	1076	12 Cns06MOG	AL405734 T3 end of
c 40	101.5	13.1	868	12 BH161951	BH161951 ENTRB12TR
c 41	101	13.0	609	10 BJ119410	BJ119410 BJ119410
c 42	101	13.0	619	9 AU180092	AU180092 AU180092
c 43	101	13.0	936	12 Cns07CJY	AL439288 T7 end of
c 44	100.5	12.9	864	12 A2547838	A2547838 ENTF21TR
c 45	99.5	12.8	522	10 B1215589	B1215589 RE22139.5

## ALIGNMENTS

RESULT 1	CNS07CG2	938 bp	DNA	linear	GSS 08-JUL-2001
CNS07CG2/c	T7 end of clone BD0AA002F07 of library BD0AA	from strain CBS 94	of		
LOCUS	Candida tropicalis	genomic survey sequence.			
DEFINITION	AL439128	GI:12222541			
ACCESSION	AL439128				
VERSION	GSS.				
KEYWORDS	Candida tropicalis.				
SOURCE	Candida tropicalis				
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.				
REFERENCE	1 (bases 1 to 938)				
AUTHORS	Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and Dujon,B.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis				
JOURNAL	FEBS Lett. 487 (1), 91-94 (2000)				
MEDLINE	20584726				
REFERENCE	2 (bases 1 to 938)				
AUTHORS	Souci�t,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingie,A., Llorente,B., Malpertuy,A., Neuviglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nloche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.				

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)  
MEDLINE 20584711  
REFERENCE 3 (bases 1 to 938)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: sequef@genoscope.cns.fr - Web: This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

## FEATURES

Location/Qualifiers

1..938

/organism="Candida tropicalis"

/strain="CBS 94"

/db\_xref="taxon:5482"

/clone="BD0AA002F07"

/clone\_lib="BD0AA"

/note="end : T7"

BASE COUNT 155 a 246 c 196 g 339 t 2 others

## ORIGIN

Alignment Scores:  
Pred. No.: 0.0021 Length: 938  
Score: 127.50 Matches: 54  
Percent Similarity: 44.38% Conservative: 21  
Best Local Similarity: 31.95% Mismatches: 59  
Query Match: 16.41% Indels: 35  
DB: 12 Gaps: 8

US-09-847-539A-6 (1-159) x CNS07CG2 (1-938)

Qy 6 GluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaPro 25  
Db 677 GAAGCACCAGAGGTAGCAGACATCGGATTCACCTAGCAGACAGATGACGACCA 618  
Qy 26 Glu-----LysLeuAlaLeuArgAsnGluGluArg----- 35  
Db 617 GAATGCGCAGATGCATTCAGTTCATCAGCAGAGCAGATGAAGCACCAGAGTGGA 558  
Qy 36 ---AlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThrAlaIle 54  
Db 557 GATGCATCGGATTCATAGCATCAGACGACATGAACACACGAGATGGCAGATCGATTG 498  
Qy 55 GluAlaAlaSerSerAspAlaLeuGluAla-----LeuAlaAspGlnThrAspAlaLeu 72  
Db 497 GATTCATCAGTACGATCAGAGAGATGAACACACGAGATGGCAGATTCGATTCACATA 438  
Qy 73 GlnSerGluGluAlaAlaValValLysAla---AspAsnAlaAlaSer----- 87  
Db 437 GCATCAGAGCACCAGAGGTAGCAGACGATCGGATTCACCGCATCAGACGATGAA 378  
Qy 88 -----AspAlaLeuGluAlaLeu----- 93  
Db 377 ACACGAGAAGTGGCAGATGCATTCGATTCATCAGATCAGACGACATGAACACACGAG 318  
Qy 94 ---AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValGln-----SerAsp 111  
Db 317 GTGCGCAGATGCATTCGATTCATCAGATCAGAGCAGACGAGATGGATTCGATTCGAT 258  
Qy 112 AsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAsp---ValLys 130  
Db 257 TCACCTAGCATCAGACGACATGAACACACGAGATGGCAGATTCGATTCATCAGCA 198

Qy 131 LysThrLysAspThrLysProValVal 139  
Db 197 TCAGAGCAGATGAACACACGACGATG 171

## RESULT 2

BF480699

LOCUS

DEFINITION

BF480699 689 bp mRNA linear EST 20-FEB-2001  
L0-2660T3 Ice plant Lambda Uni-Zap XR expression library, 0 hours  
NaCl treatment Mesembryanthemum crystallinum cDNA clone L0-2660 5',  
mRNA sequence.

ACCESSION

BF480699

VERSION

BF480699.1 GI:11551526

KEYWORDS

EST.

SOURCE

ORGANISM

Mesembryanthemum crystallinum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Caryophyllidae; Caryophyllales; Alzooceae; Mesembryanthemum.

REFERENCE

1 (bases 1 to 689)

AUTHORS

Cushman, J.C.

TITLE

An expressed sequence tag database for the common ice plant,  
Mesembryanthemum crystallinum

JOURNAL

Unpublished (1997)

COMMENT

Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200 Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu  
PCR Primers  
FORWARD: T7  
BACKWARD: T3  
Plate: L0-27 row: E column: 12  
Seq primer: T3  
High quality sequence stop: 350  
POLYA-No.

## FEATURES

Location/Qualifiers

1..689

/organism="Mesembryanthemum crystallinum"

/db\_xref="taxon:3544"

/clone="L0-2660"

/clone\_lib="Ice plant Lambda Uni-Zap XR expression library  
0 hours NaCl treatment"

/tissue\_type="Leaf"

/dev\_stage="Six week old"

/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"

BASE COUNT 203 a 160 c 186 g 140 t

## ORIGIN

Alignment Scores:  
Pred. No.: 0.0509 Length: 689  
Score: 112.50 Matches: 47  
Percent Similarity: 40.74% Conservative: 19  
Best Local Similarity: 29.01% Mismatches: 53  
Query Match: 14.48% Indels: 43  
DB: 10 Gaps: 8

US-09-847-539A-6 (1-159) x BF480699 (1-689)

Qy 20 LeuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGlu 39  
Db 226 ATTATGGGGACTGTGCGNACAAG---CCGTTGACGACGAGAGGGGCG----- 273  
Qy 40 LeuLysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaSerSer 59  
Db 274 -----GCGCGCGCGCGCGAGGAGGAGGAGCA-----GCGGTGGAAGTTGCGCGCG 321  
Qy 60 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAla-----LeuGlnSerGlu 75  
Db 322 GAAGGTGCGGAAGTGTGCGGAAAGTAGAGATATTGCGGTGTGTGACTTAAAAAGGAG 381



This clone was isolated from Medaka liver cDNA library (OLE).

```
FEATURES
  source
    1..442
      /organism="Oryzias latipes"
      /strain="HNI"
      /db_xref="taxon:8090"
      /clone="OLE18.01d"
      /clone_lib="Medaka liver cDNA library (OLE) from HNI"
      /tissue_type="liver"
      /dev_stage="adult"
BASE COUNT  127 a 169 c 52 g 94 t
ORIGIN
Alignment Scores:
  Pred. No.:      0.0339      Length:      442
  Score:          112.00      Matches:      39
  Percent Similarity: 45.67%   Conservative: 19
  Best Local Similarity: 30.71% Mismatches:      55
  Query Match:      14.41%   Indels:       14
  DB:               9       Gaps:         4
US-09-847-539A-6 (1-159) x AU180004 (1-442)
QY  24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGln 43
Db  390 GCTGAGGATCTGAGGTTTGAAGGCTACTGAGGCTGCTGAGGATCTGAGGATCTGAG 331
QY  44 AlaIleGluAspLysGluAlaThrAlaIleGluAlaA---SerSerAspAlaLeu 62
Db  330 GCTGCTGAGGATCTGAGGTTTGAAGGATCTGAGGCTGCTGAGGATCTGAGGATGTT 271
QY  63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAla 82
Db  270 GAGGCT---GCTGAGGATCTGAGGTTTGAAGGATCTGAGGCTGCTGAGGATCTGAG 214
QY  83 AspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102
Db  213 GATACTGAGGCTGATGAGGCTGCTGAGGTTTGAAGGATCTGAGGCTGCTGAGGATCT 154
QY  103 GluGlu-----AlaGluValValGlnSerAspAsnAlaAlaSerAsp 116
Db  153 GAGGATACTGAGGCTGATGAGGCTGCTGAGGTTTGAAGGATCTGAGGCTGCTGAGGAT 94
QY  117 AlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThrLys 136
Db  93 -----TCTGAGGTTTGAAGGATCTGAGGCTGCTGAGGATCTGAGGATCTGAG 52
QY  137 ProValValLysLysGluGlu 143
Db  51 GCTGTTGAGGCTGCTGAGGAT 31
RESULT 5
BM029212 475 bp mRNA linear EST 05-NOV-2001
LOCUS IpSk000051 Skin cDNA library Ictalurus punctatus cDNA 5', mRNA
DEFINITION
ACCESSION BM029212
VERSION BM029212
KEYWORDS EST.
SOURCE channel catfish.
ORGANISM Ictalurus punctatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
Ictaluridae; Ictalurus.
1 (bases 1 to 475)
REFERENCE Karsi,A., Cao,D., Li,P., Ju,Z., Kocbas,A., Feng,J., Patterson,A.,
AUTHORS Mickett,K.D. and Liu,Z.
TITLE Transcriptome analysis of channel catfish (Ictalurus punctatus):
Initial analysis of gene expression and microsatellite-containing
cDNAs in the skin
JOURNAL Unpublished (2001)
COMMENT Contact: Liu ZJ
The Fish Molecular Genetics and Biotechnology Laboratory.
```

Department of Fisheries and Allied Aquacultures and Program of Cell  
and Molecular Biosciences  
Auburn University  
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA  
Tel: 334 844 4054  
Fax: 334 844 9208  
Email: zliuecesag.auburn.edu  
Seq primer: M13 Reverse.

FEATURES  
source  
1..475  
/organism="Ictalurus punctatus"  
/db\_xref="taxon:7998"  
/clone\_lib="Skin cDNA library"  
/note="Organ: Skin; Vector: pSport1; Site\_1: NotI; Site\_2:  
Sall"

BASE COUNT 229 a 77 c 154 g 15 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.0369 Length: 475  
Score: 112.00 Matches: 37  
Percent Similarity: 53.72% Conservative: 28  
Best Local Similarity: 30.58% Mismatches: 45  
Query Match: 14.41% Indels: 11  
DB: 10 Gaps: 3  
US-09-847-539A-6 (1-159) x BM029212 (1-475)  
QY 29 AlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLys 48  
Db 91 GCAGCAGCAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCA-----GAAGAA 144  
QY 49 GluAlaThrAlaIleGluAlaAlaLeuSerSerAspAlaLeuGluAlaLeuAlaLeuGln 68  
Db 145 GAAGCAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCAAGCA 204  
QY 69 ThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAsp 88  
Db 205 GAAGAAGAAAGCAGCAGCAGCAGCAGAA-----GAAGCAGAAAGAAAGAAAGAA 258  
QY 89 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGlu----- 106  
Db 259 GCAGCAGAAAGCAGCAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCAAGCAAGCA 318  
QY 107 -----ValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAla 122  
Db 319 GCAGAAGAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAAGCA 378  
QY 123 ThrProIleAlaLeuAspVal-LysLysThrLysAspThrLysProValValLysLysGln 142  
Db 379 GAAGCAGCAGCAGAAAGAAAGCAGCAGCAGCAGCAGCAGCAGAAAGAAAGCAAGCA 438  
QY 142 u 142  
Db 439 T 439  
RESULT 6  
AUI79963 570 bp mRNA linear EST 21-MAR-2001  
LOCUS AUI79963/c  
DEFINITION cDNA clone OLe17.07a similar to gb|AB025967| Oryzias latipes mRNA  
for chorionogenin HmlnR, complete cds, mRNA sequence.  
ACCESSION AUI79963  
VERSION AUI79963.1 GI:13428800  
KEYWORDS EST.  
SOURCE Japanese medaka.  
ORGANISM Oryzias latipes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.  
1 (bases 1 to 570)  
REFERENCE Naruse,K., Mitani,H. and Tanaka,M.  
AUTHORS

TITLE Medaka EST Project in University of Tokyo (2001)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Kiyoshi Naruse  
Department of Biological Sciences  
Graduate School of Science, University of Tokyo  
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan  
Tel: 81-3-5841-4443  
Fax: 81-3-5841-4410  
Email: naruse@biol.s.u-tokyo.ac.jp  
This clone was isolated from Medaka liver cDNA library (OLE).

FEATURES  
source  
1..570  
/organism="Oryzias latipes"  
/strain="HNI"  
/db\_xref="taxon:8090"  
/clone="OLE17.07a"  
/clone\_lib="Medaka liver cDNA library (OLE) from HNI"  
/tissue\_type="liver"  
/dev\_stage="adult"  
BASE COUNT 157 a 203 c 84 g 126 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.0458 Length: 570  
Score: 112.00 Matches: 39  
Percent Similarity: 45.67% Conservative: 19  
Best Local Similarity: 30.71% Mismatches: 55  
Query Match: 14.41% Indels: 14  
DB: 9 Gaps: 4

US-09-847-539A-6 (1-159) x AU179963 (1-570)

Qy 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGln 43  
Db 494 GCTGAGGATCTCTGAGGTTTGAAGGCTACTGAGGCTCTGAGGATCTGAGGATGTTGAG 435

Qy 44 AlaIleGluAspLysGluAlaThrAlaIleGluAlaA---SerSerAspAlaLeu 62  
Db 434 GCTGCTGAGGATCTGAGGTTTGAAGGATCTGAGGCTCTGAGGATCTGAGGATGTT 375

Qy 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 82  
Db 374 GAGGCT---GCTGAGGATCTGAGGTTTGAAGGATCTGAGGCTCTGAGGATCTGAG 318

Qy 83 AspAsnAlaAlaSerAspAlaLeuGluAlaAlaAspGlnThrAspAlaLeuGlnSer 102  
Db 317 GATACTGAGGCTGATGAGGCTGCTGAGGTTTGAAGGATCTGAGGCTGCTGAGGATCT 258

Qy 103 GluGlu---AlaGluValValGlnSerAspAsnAlaAlaSerAsp 116  
Db 257 GAGGATCTGAGGCTGATGAGGCTGCTGAGGTTTGAAGGATCTGAGGCTGCTGAGGAT 198

Qy 117 AlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLys 136  
Db 197 -----TCTGAGGTTTGAAGGATCTGAGGCTGCTGAGGATCTGAG 156

Qy 137 ProValValLysLysGluGlu 143  
Db 155 GCTGTTGAGGCTGCTGAGGAT 135

RESULT 7  
AI057981/c 615 bp mRNA linear EST 20-JUL-1998  
LOCUS SHOWL3CAN21G06SK Onchocerca volvulus infective larva cDNA  
DEFINITION (SAW94WL-OVL3) Onchocerca volvulus cDNA clone SMOVL3CAN21G06 5', mRNA sequence.  
ACCESSION AI057981  
VERSION AI057981.1 GI:3331847  
KEYWORDS EST.  
SOURCE Onchocerca volvulus.  
ORIGIN Onchocerca volvulus.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
Onchocercidae; Onchocerca.

# REFERENCE

1 (bases 1 to 615)  
Williams,S.A., Lu,W., Lizotte-Waniewski,M. and Laney,S.J.  
Genes expressed in infective third stage larvae of Onchocerca  
volvulus

JOURNAL Unpublished (1995)  
COMMENT Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135833826  
Fax: 4135833786  
Email: genome@smith.edu  
Seq primer: pBluescript SK.

# FEATURES

source  
1..615  
Location/Qualifiers

/organism="Onchocerca volvulus"  
/strain="Sierra Leone"  
/db\_xref="taxon:6282"  
/clone="SMOVL3CAN21G06"  
/clone\_lib="Onchocerca volvulus infective larva cDNA  
(SAW94WL-OVL3)"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: lambda UniZap XR; Site1: EcoR I; Site2:  
Xho I; Cutaneous filarial nematode parasite of humans.  
mRNA was prepared from third stage infective larvae of  
Onchocerca volvulus isolated from mosquitoes 10 days after  
infection and converted to double stranded cDNA using  
reverse transcriptase and oligo(dT) followed by RNase H  
and DNAPol I. The library had 1.8 x 10E5 independent  
recombinants and average insert size was 900 base pairs.  
The library was constructed by Wenhong Lu. The library is  
available from Dr. S.A. Williams, email genome@smith.edu."

BASE COUNT 74 a 186 c 176 g 173 t 6 others  
ORIGIN

# Alignment Scores:

Pred. No.: 0.0564 Length: 615  
Score: 111.50 Matches: 38  
Percent Similarity: 40.29% Conservative: 18  
Best Local Similarity: 27.34% Mismatches: 48  
Query Match: 14.35% Indels: 35  
DB: 9 Gaps: 3

US-09-847-539A-6 (1-159) x AI057981 (1-615)

Qy 44 AlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAspAlaLeuGlu 63  
Db 552 GCATNGTCAGGTCAGGAAGGCTCNTCCGGCGCAGAGCGGCATCAGCAAGGCCACTGAA 493

Qy 63 -----  
Db 492 GCGGAAAAAGTGC CGCAGCGCAGAGTCCTCAAAAAACGCGCGGCCACACGTCGCGGT 433

Qy 64 ---AlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 82  
Db 432 CGCGCGAAACGCTCAGAACGATGCTGCACGCTCACAACATCAGCGCCACGCTCTGCN 373

Qy 83 AspAsnAlaAlaSerAspAlaLeuGluAlaAlaAspGlnThrAspAlaLeuGlnSer 102  
Db 372 TCCACCGCGCGCAGAAAGCGTCAGAGCGCCCTTCAGCACGAGATCGCGTGGCCTCA 313

Qy 103 GluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAla 122  
Db 312 AAGAGGCGCAAAATCATCAGAACGACGATCATCAAGTGCCTGCGTGCAGCTTCC 253

Qy 123 ThrProIleAlaLeuAsp-----ValLysLysThrLysAspThrLysProVal 138  
Db 252 TCGGCAACGGCGCGAGAAATTTCTGCCAGCGCGCGCAAAAGCTCCGAGACG----- 202

Qy 139 ValLysLysGluGluArgGlnAsnValAsnThrLeuProThrGlyGluGluSer 157  
Db 201 -----AATGCCAGGTCATCTGAAACAGCAGCGGACGAGC 166

```

RESULT 8
AUI80137/c
LOCUS      AUI80137 Medaka liver cDNA library (OLE) from HNI Oryzias latipes
DEFINITION CDNA clone OLe19.11e similar to gb|AB025967| Oryzias latipes mRNA
            for chorionogenin Hminor, complete cds, mRNA sequence.
ACCESSION  AUI80137
VERSION     AUI80137.1 GI:13428974
KEYWORDS   EST.
SOURCE     Japanese medaka.
ORGANISM   Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE  1 (bases 1 to 344)
AUTHORS   Naruse,K., Mitan,H. and Tanaka,M.
TITLE     Medaka EST Project in University of Tokyo (2001)
JOURNAL   Unpublished (2001)
COMMENT   Contact: Kiyoshi Naruse
            Department of Biological Sciences
            Graduate School of Science, University of Tokyo
            Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
            Tel: 81-3-5841-4443
            Fax: 81-3-5841-4410
            Email: naruse@biol.s.u-tokyo.ac.jp
            This clone was isolated from Medaka liver cDNA library (OLE).

FEATURES             source
BASE COUNT           96 a 132 c 41 g 74 t 1 others
ORIGIN
1..344
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone="OLE19.11e"
/clone_lib="Medaka liver cDNA library (OLE) from HNI"
/tissue_type="liver"
/dev_stage="adult"

Alignment Scores:
Pred. No.:      0.0319      Length:      344
Score:          111.00      Matches:      38
Percent Similarity: 46.49%      Conservative: 15
Best Local Similarity: 33.33%      Mismatches: 53
Query Match:     14.23%      Indels:      8
DB:              9          Gaps:      3

US-09-847-539A-6 (1-159) x AUI80137 (1-344)
QY  24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysGln 43
    ||| ::: |||::: ||| ||| |||
DB  330 GCTGAGGATTNTGAGGTTTGAAGGCTACTGAGGCTGCTGAGGATCTCGAGGATGTTGAG 271
    ||| |||::: |||::: ||| ||| |||
QY  44 AlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAala---SerSerAspAlaLeu 62
    ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  270 GCTGCTGAGGATCTCGAGGCTTTGAAGGATCTGAGGCTGCTGAGGATCTCGAGGATGTT 211
    ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 82
    ||| ||| |||::: |||::: |||::: ||| ||| |||
DB  210 GAGGCT---GCTGAGGATCTGAGGTTTGAAGGATCTGAGGCTGCTGAGGATCTGAG 154
    ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  83 AspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102
    ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  153 GATAGTGGCTGATGAGGCTGCTGAGGTTTGAAGGATCTGAGGCTGCTGAGGATCTCT 94
    ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  103 GluGluAlaGluAlaValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAla 122
    |||::: ||| ::::: |||
DB  93 GAGGATAGTGGCTGATGAGGCTGCTGAGGTTTGAAGGAT----- 52
    ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  123 ThrProIleAlaLeuAspValLysLysThrLysAspThrLys 136
    ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  51 ACTGAGGCTGCTGAGGATCTCGAGGTTTGAAGGATCTGAG 10
    ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 9
BM355271
LOCUS      BM355271
DEFINITION r12zh07.yl Globodera rostochiensis J2 pcDNAII Smant v1 Globodera
            rostochiensis cDNA 5', mRNA sequence.
ACCESSION  BM355271
VERSION     BM355271.1 GI:18089902
KEYWORDS   EST.
SOURCE     Globodera rostochiensis.
ORGANISM   Globodera rostochiensis
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida;
            Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
REFERENCE  1 (bases 1 to 617)
AUTHORS   McCarter,J., Clifton,S., Chlapelli,B., Pape,D., Martin,J., Wylie,T.,
            Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
            Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagaris,V., R.,
            Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Stepcoe
            ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
            , Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
TITLE     The Washington Univ. Nematode EST Project, 1999
JOURNAL   Unpublished (1999)
COMMENT   Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was contributed by Dr. Geert Smant of the Laboratory of
            Nematology at Wageningen University, Wageningen,
            Netherlands(geert.smant@nema.dpw.wau.nl). DNA Sequencing by:
            Washington University Genome Sequencing Center
            Seq primer: -40RP from Gibco
            High quality sequence stop: 483.
            Location/Qualifiers
                1..617
                /organism="Globodera rostochiensis"
                /db_xref="taxon:31243"
                /clone_lib="Globodera rostochiensis J2 pcDNAII Smant v1"
                /dev_stage="J2"
                /lab_host="DH10B"
                /note="Vector: pcDNAII (Invitrogen); Site 1: BstXI;
                Site 2: EcoRI; The library was donated for sequencing by
                Geert Smant from Wageningen University, Laboratory of
                Nematology, The Netherlands."
BASE COUNT      207 a 178 c 122 g 110 t
ORIGIN
1..617
Alignment Scores:
Pred. No.:      0.0718      Length:      617
Score:          110.50      Matches:      36
Percent Similarity: 47.10%      Conservative: 29
Best Local Similarity: 26.09%      Mismatches: 54
Query Match:     14.22%      Indels:      19
DB:              10          Gaps:      4

US-09-847-539A-6 (1-159) x BM355271 (1-617)
QY  32 AsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLys----- 48
    ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  208 ATGAGAGTCTCTCAACAGACGAGCATCATCCCCGCCACTGAGAGTCTCAACAGAC 267
    ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  49 -----GluAlaThrAlaIleGluAlaAlaSerSerAspAla 61
    ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  268 GAAGCATCATCCCGGCCACTGAGGAGTCTCAGCATCGGAGCATCATCTCGGCTCT 327
    ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 81
    ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  328 GAAGAGTCTCAACAGACGAGCATCATCTACGGGTACTCAACAGATCTTCAACAGACAA 387
    ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101
    ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

Db 366 TTGAAGGATACTAGGCTGCTGAGGCTGCTGAGGCTGCTTAGGATTCTGAGGCTGCTGAG 307
Qy 64 AlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla--- 82
Db 306 GCTGCTTAGGATTCTGAGGATACTAGGCTGCTGATCACTTCTGATGTTTGAAGGCTACT 247
Qy 83 -----AspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 98
Db 246 GAGGCTGTTTAGGATTCTGAGGATACTAGGCTGCTGATGAGGCTCTTAGGATTCTGAGGAT 187
Qy 99 AlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTnp 118
Db 186 ACTGAGCGGCTGAGGATTCTGAGGTTTGAAGGCTACTGAGGCTGCTGAGGATTCTGAG 127
Qy 119 GluLysAlaAlaThrProLeuAlaLeuAspValLysLysThrLysAspThrLysProval 138
Db 126 GATGTTGAGGCT-----GCTGAGGATTCTGAGGTTTGAAGGATACTGAGGCTGCT 76
Qy 139 ValLysLysGluGlu 143
Db 75 GAGGATTCTGAGGAT 61

RESULT 12
AUI80168/c
LOCUS
DEFINITION
AUI80168 Medaka liver cDNA library (OLE) from HNI Oryzias latipes
cDNA clone OLE20.04g similar to pIR148048| egg envelope protein wf
- winter flounder, mRNA sequence.
ACCESSION
AUI80168
VERSION
AUI80168.1 GI:13429005
KEYWORDS
EST.
SOURCE
Japanese medaka.
ORGANISM
Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE
1 (bases 1 to 610)
Naruse,K., Mitani,H. and Tanaka,M.
Medaka EST Project in University of Tokyo (2001)
Unpublished (2001)
Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from Medaka liver cDNA library (OLE).
FEATURES
source
1..610
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone="OLE20.04g"
/clone_lib="Medaka liver cDNA library (OLE) from HNI"
/tissue_type="liver"
/dev_stage="adult"
BASE COUNT 173 a 189 c 104 g 144 t
ORIGIN
1..610
Alignment Scores:
Pred. No.: 0.0797 Length: 610
Score: 110.00 Matches: 40
Percent Similarity: 46.90% Conservative: 28
Best Local Similarity: 27.59% Mismatches: 55
Query Match: 14.16% Indels: 22
DB: 9 Gaps: 4

US-09-847-539A-6 (1-159) x AUI80168 (1-610)
Qy 18 ThrAsnLeuLeuGlyAsnAlaProGluLysLeuAlaLeuArg----- 31

```

```

Db 438 AGCAGTTTATGTCACACTCCGAGCGCAGTGTGTTTGAGGAGCCACAAGGTATTTTAT 379
Qy 32 -----AsnGluGluArgAlaAlaLeuLeuLysLysGlnAla 44
Db 378 AATACATCATCTACTCACAAGAGTGAATTTGATGGTGGAGTCTGAGGATT 319
Qy 45 IleGluAspLysGluAlaThrAlaAlaIleGluAlaAla---SerSerAspAlaLeuGlu 63
Db 318 TTGAAGGATACTAGGCTGCTGAGGCTGCTGAGGCTCTTAGGATTCTGAGGCTGCTGAG 259
Qy 64 AlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla--- 82
Db 258 GCTGCTTAGGATTCTGAGGATACTGAGGCTGCTGATCATCTGATCTGTTTGAAGGTTACT 199
Qy 83 -----AspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 98
Db 198 GAGGCTGTTTAGGATTCTGAGGATACTGAGGCTGATGAGGCTGCTTAGGATTCTGAGGAT 139
Qy 99 AlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTnp 118
Db 138 ACTGAGCGGCTGAGGATTCTGAGGTTTGAAGGCTACTGAGGCTGCTGAGGATTCTGAG 79
Qy 119 GluLysAlaAlaThrProLeuAlaLeuAspValLysLysThrLysAspThrLysProval 138
Db 78 GATGTTGAGGCT-----GCTGAGGATTCTGAGGTTTGAAGGATACTGAGGCTGCT 28
Qy 139 ValLysLysGluGlu 143
Db 27 GAGGATTCTGAGGAT 13

RESULT 13
AUI79269/c
LOCUS
DEFINITION
AUI79269 Medaka liver cDNA library (OLE) from HNI Oryzias latipes
cDNA clone OLE06.03e similar to gb|AB025967| Oryzias latipes mRNA
for choriolegen in Hminor, complete cds, mRNA sequence.
ACCESSION
AUI79269
VERSION
AUI79269.1 GI:13428106
KEYWORDS
EST.
SOURCE
Japanese medaka.
ORGANISM
Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE
1 (bases 1 to 453)
Naruse,K., Mitani,H. and Tanaka,M.
Medaka EST Project in University of Tokyo (2001)
Unpublished (2001)
Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from Medaka liver cDNA library (OLE).
FEATURES
source
1..453
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone="OLE06.03e"
/clone_lib="Medaka liver cDNA library (OLE) from HNI"
/tissue_type="liver"
/dev_stage="adult"
BASE COUNT 137 a 159 c 59 g 98 t
ORIGIN
1..453
Location/Qualifiers
Alignment Scores:
Pred. No.: 0.0631 Length: 453
Score: 109.50 Matches: 41

```







```

source
1..664
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0022N09f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/notes="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 221 a 124 c 177 g 141 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 0.179 Length: 664
Score: 107.00 Matches: 39
Percent Similarity: 44.97% Conservative: 28
Best Local Similarity: 26.17% Mismatches: 58
Query Match: 13.77% Indels: 24
DB: 10 Gaps: 5

US-09-847-539A-6 (1-159) x BG443972 (1-664)
QY 9 ArgIleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeu 28
Db 23 CAATTATTTCCCACTTCCTTTTCCCTTCAAAATTTCTTGGGATTCCTATGGCTACTGCT 82
QY 29 AlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys---LysGlnAlaIleGluAsp 47
Db 83 GAGGTTGCATCATGTTCAAACTGCACCTTCATGAGGAGAAACTGAAGATCAGTGAAGGAT 142
QY 48 LysGluAlaThrThrAlaIleGluAlaIleSerSerAspAlaLeuGluAlaLeuAlaAsp 67
Db 143 CAAGAAGCTACACAGATGAAGTAGTTGCGGAAGCTTCGCGGAGGAGGAGGCTGCCGAA 202
QY 68 GlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaIleSer 87
Db 203 GAGCCAAAG-----GAAGCGGAGCTGCGGCGGTGCGAGTTCGAGGAACCTGAAGCCTCG 256
QY 88 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluVal 107
Db 257 -----GTTGAAGTTGAAACCAAGGAGGTG 280
QY 108 ValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeu 127
Db 281 GTAGAAGAACCAAGGCTGTACTGAG-----GAGCCAGTAGTAGAG 322
QY 128 AspValLysLysThrLysAspThrLysProVal-----ValLysLysGluGluArg 144
Db 323 GAGACTCTAAGGAACAGTCGAGAACCAACCTGCGGAGGAGATCAAGGAACCACTGAA 382
QY 145 GlnAsnValAsnThrLeuProThrThr 153
Db 383 CAACACGTTGAACCAAGAACTACA 409

RESULT 18
BM360707 857 bp mRNA linear EST 09-JAN-2002
LOCUS GA_Ea0034C23r Gossypium arboreum 7-10 dpa fiber library Gossypium
DEFINITION arboreum cDNA clone GA_Ea0034C23r, mRNA sequence.
ACCESSION BM360707
VERSION
KEYWORDS
SOURCE
ORGANISM
Gossypium arboreum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 857)
REFERENCE
Wing,R.A., Friisch,D., Yu,Y., Mein,D., Rambo,T., Simmons,J., Henry
D., Wood,T.C., Leslie,A. and Wilkins,T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total High Quality bases = 618
Seq primer: TAATACGACTCACTATAGG
High quality sequence stop: 846.
FEATURES
Location/Qualifiers
1..857
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0034C23r"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/notes="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 271 a 150 c 230 g 206 t
ORIGIN

Alignment Scores:
Pred. No.: 0.243 Length: 857
Score: 107.00 Matches: 39
Percent Similarity: 44.97% Conservative: 28
Best Local Similarity: 26.17% Mismatches: 58
Query Match: 13.77% Indels: 24
DB: 10 Gaps: 5

US-09-847-539A-6 (1-159) x BM360707 (1-857)
QY 9 ArgIleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeu 28
Db 13 CAATTATTTCCCACTTCCTTTTCCCTTCAAAATTTCTTGGGATTCCTATGGCTACTGCT 72
QY 29 AlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys---LysGlnAlaIleGluAsp 47
Db 73 GAGGTTGCATCATGTTCAAACTGCACCTTCATGAGGAGAAACTGAAGATCAGTGAAGGAT 132
QY 48 LysGluAlaThrThrAlaIleGluAlaIleSerSerAspAlaLeuGluAlaLeuAlaAsp 67
Db 133 CAAGAAGCTACACAGATGAAGTAGTTGCGGAAGCTTCGCGGAGGAGGAGGCTGCCGAA 192
QY 68 GlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaIleSer 87
Db 193 GAGCCAAAG-----GAAGCGGAGCTGCGGCGGTGCGAGTTCGAGGAACCTGAAGCCTCG 246
QY 88 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluVal 107
Db 247 -----GTTGAAGTTGAACCAAGGAGGTG 270
QY 108 ValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeu 127
Db 271 GTAGAAGAACCAAGGCTGTACTGAG-----GAGCCAGTAGTAGAG 312
QY 128 AspValLysLysThrLysAspThrLysProVal-----ValLysLysGluGluArg 144
Db 313 GAGACTCTAAGGAACAGTCGAGAACCAACCTGCGGAGGAGATCAAGGAACCACTGAA 372
QY 145 GlnAsnValAsnThrLeuProThrThr 153
Db 373 CAACACGTTGAACCAAGAACTACA 399

RESULT 19
BG445649 881 bp mRNA linear EST 15-MAR-2001
LOCUS GA_Ea0029A12f Gossypium arboreum 7-10 dpa fiber library Gossypium
DEFINITION

```

arborescence cDNA clone GA\_Ea029A12f, mRNA sequence.

```
ACCESSION      BG445649
VERSION        BG445649.1 GI:13355301
KEYWORDS       EST.
SOURCE         Gossypium arboreum.
ORGANISM       Gossypium arboreum.
REFERENCE      1 (bases 1 to 881)
AUTHORS        Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
               D., Wood, T.C., Leslie, A. and Wilkins, T.A.
TITLE          An integrated analysis of the genetics, development, and evolution
               of the cotton fiber
JOURNAL        Unpublished (2000)
COMMENT        Contact: Wing RA
               Clemson University Genomics Institute
               100 Jordan Hall, Clemson, SC 29634, USA
               Tel: 864 656 7288
               Fax: 864 656 4293
               Email: rwing@clemson.edu
               Seq primer: TAATACGACTCCTACTAGGG
               High quality sequence start: 10
               High quality sequence stop: 715.
               Location/Qualifiers
FEATURES       source
               1..881
               /organism="Gossypium arboreum"
               /strain="AKA"
               /cultivar="8400"
               /db_xref="taxon:29729"
               /clone="GA_Ea029A12f"
               /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
               /tissue_type="Fibers isolated from bolls harvested 7-10
               dpa"
               /lab_host="E. coli"
               /note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT    275 a 162 c 237 g 205 t 2 others
ORIGIN
Alignment Scores:
Pred. No.:      0.251      Length:      881
Score:          107.00     Matches:      39
Percent Similarity: 44.97%   Conservative: 28
Best Local Similarity: 26.17% Mismatches: 58
Query Match:    13.77%     Indels:      24
DB:             10        Gaps:       5

US-09-847-539A-6 (1-159) x BG445649 (1-881)
Qy  9 ArgilleleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeu 28
Db  30 CAATATTTCCTCCACTTCCTTTGGCTTCAAAATTTCTTGGGATTCCTACTGCTACTGCT 89
Qy  29 AlaLeuArgAsnGluGluArgAlaLeuAspGluLeuLys---LysGlnAlaIleGluAsp 47
Db  90 GAGGTTGCATCAGTTCAAACCTGCCTTCATGAGGAGAAACCTGAAGAATCAGTGAAGGAT 149
Qy  48 LysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAsp 67
Db  150 CARGAAGCTACACAGATGAAGTAGTTCCTCCGAGGCTCTCTGCGGCGGAGCTGCCCAA 209
Qy  68 GlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAlaAlaSer 87
Db  210 GAGCCAAAG-----GAGCGGAGCCTCGCGGGTGGCGAGCTCGGAGACCTGAAGCCTCG 263
Qy  88 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluVal 107
Db  264 -----GTTGAAGTTGAACCAACCAAGGAGGTG 287
Qy  108 ValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeu 127
Db  288 CTAGAAGAACCAAGGCTGTTACTGAG-----GAGCCAGTAGTAGAG 329
```

```
Qy  128 AspValLysLysThrLysAspThrLysProVal-----ValLysLysGluGluArg 144
Db  330 GAGACTCCTAAGGAACACAGTCCAGAACCACTTTCGCGAGGAGATCAAGGAACCACTGAA 389
Qy  145 GlnAsnValAsnThrLeuProThrThr 153
Db  350 CAACCGTGTGAACCAACCAAGAACTACA 416
RESULT 20
BG444821
LOCUS        BG444821
DEFINITION  GA_Ea0025L07f Gossypium arboreum 7-10 dpa fiber library Gossypium
arborescence cDNA clone GA_Ea0025L07f, mRNA sequence.
ACCESSION    BG444821
VERSION      BG444821.1 GI:13354473
KEYWORDS     EST.
SOURCE       Gossypium arboreum.
ORGANISM     Gossypium arboreum.
REFERENCE    1 (bases 1 to 887)
AUTHORS      Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
               D., Wood, T.C., Leslie, A. and Wilkins, T.A.
TITLE        An integrated analysis of the genetics, development, and evolution
               of the cotton fiber
JOURNAL      Unpublished (2000)
COMMENT      Contact: Wing RA
               Clemson University Genomics Institute
               100 Jordan Hall, Clemson, SC 29634, USA
               Tel: 864 656 7288
               Fax: 864 656 4293
               Email: rwing@clemson.edu
               Seq primer: TAATACGACTCCTACTAGGG
               High quality sequence stop: 664.
               Location/Qualifiers
FEATURES     source
               1..887
               /organism="Gossypium arboreum"
               /strain="AKA"
               /cultivar="8400"
               /db_xref="taxon:29729"
               /clone="GA_Ea0025L07f"
               /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
               /tissue_type="Fibers isolated from bolls harvested 7-10
               dpa"
               /lab_host="E. coli"
               /note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT  272 a 161 c 241 g 213 t
ORIGIN
Alignment Scores:
Pred. No.:      0.253      Length:      887
Score:          107.00     Matches:      39
Percent Similarity: 44.97%   Conservative: 28
Best Local Similarity: 26.17% Mismatches: 58
Query Match:    13.77%     Indels:      24
DB:             10        Gaps:       5

US-09-847-539A-6 (1-159) x BG444821 (1-887)
Qy  9 ArgilleleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeu 28
Db  22 CAATATTTCCTCCACTTCCTTTGGCTTCAAAATTTCTTGGGATTCCTACTGCTACTGCT 81
Qy  29 AlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys---LysGlnAlaIleGluAsp 47
Db  82 GAGGTTGCATCAGTTCAAACCTGCCTTCATGAGGAGAAACCTGAAGAATCAGTGAAGGAT 141
Qy  48 LysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAsp 67
Db  142 CAAGAAGCTACACAGATGAAGTAGTTCCTCCGAGGCTCTCTGCGGCGGAGCTGCCCAA 201
```





```

Qy 58 rSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAl 78
      :::::||||| ||||| ||||| ::::: ||||| ||||| ::::: ||||| |||||
Db 610 GACGAGAGCCGACAGCAGAGGCGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGCAGCT 669
      :::::||||| ||||| ||||| ::::: ||||| ||||| ::::: ||||| |||||
Qy 78 aValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAs 98
      :::::||||| ||||| ||||| ::::: ||||| ||||| ::::: ||||| |||||
Db 670 AGCAGCAGAGCAGCAGAGGCGAGCT---GAAGCGGAGGAGCTCACTGGCA-----ACTGA 720
      :::::||||| ||||| ||||| ::::: ||||| ||||| ::::: ||||| |||||
Qy 98 pAlaLeuGlnSerGluAlaGluAlaValValGlnSerAspAsnAla 113
      ||||| :::::||||| ||||| ||||| ::::: ||||| |||||
Db 721 TGCCGAGAGCAGAGGCGAGGCGGAGGCGGAGCAGCAGCAGTAGAGCA 766

RESULT 24
BJ001775          653 bp mRNA linear EST 05-DEC-2001
LOCUS             BJ001775 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA027A11 5',
DEFINITION        mRNA sequence.
ACCESSION         BJ001775
VERSION           BJ001775.1 GI:17364666
KEYWORDS          EST.
SOURCE            Japanese medaka.
ORGANISM          Oryzias latipes
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                  Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
                  Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE         1 (bases 1 to 653)
AUTHORS           Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE            Medaka EST Project in Takeda's lab
JOURNAL           Unpublished (2001)
COMMENT          Contact: Tadasu Shin-i
                  National Institute of Genetics
                  1111 Yata, Mishima, Shizuoka 411-8540, Japan
                  Tel: 81-559-81-6856
                  Fax: 81-559-81-6855
                  Email: tshini@genes.nig.ac.jp.
FEATURES          Location/Qualifiers
source            1..653
                  /organism="Oryzias latipes"
                  /strain="Hd-r"
                  /db_xref="taxon:8090"
                  /clone_lib="MF01SSA027A11"
                  /clone_lib="MF01SSA cDNA"
                  /sex="mixture of female and male"
                  /tissue_type="whole embryo"
                  /dev_stage="segmentation stage 20 - 25"
BASE COUNT        226 a 136 c 173 g 118 t
ORIGIN
Alignment Scores:
Pred. No.:        0.251 Length: 653
Score:            105.50 Matches: 34
Percent Similarity: 44.97% Conservative: 42
Best Local Similarity: 20.12% Mismatches: 74
Query Match:      13.58% Indels: 19
DB:               10 Gaps: 5

US-09-847-539A-6 (1-159) x BJ001775 (1-653)

Qy 2 AspSerProLeuGluGlnProArgileIleProAsnGlyGlyThrLeuThrAsnLeuLeu 21
      :::::||||| ||||| ||||| ::::: ||||| |||||
Db 141 GATGCTGGGACAAAAGAGCCGCTTTACAGAGCAGGTGGAACAGCTGACCTTCAGCT 200
      :::::||||| ||||| ||||| ::::: ||||| |||||
Qy 22 GlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys 41
      :::::||||| ||||| ||||| ::::: ||||| |||||
Db 201 GATTCATCTGACGTGAATGCTATCATGGGACAACTCT-----GAGGACCAACAGC 251
      :::::||||| ||||| ||||| ::::: ||||| |||||
Qy 42 LysGlnAlaIleGluAspLysLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAla 61
      :::::||||| ||||| ||||| ::::: ||||| |||||
Db 252 GATCATCTATGATGGACAAATCTGAGACACACCTCAAGAGCTCCACTGAACATGATGAA 311
      :::::||||| ||||| ||||| ::::: ||||| |||||
Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 81

```

```

Db 312 GAAAAGCAGAAAATGACCAAACTGACACACACAGAGGAGGAGGATTCACGATCAAGAAC 371
      :::::||||| ||||| ||||| ::::: ||||| ||||| ::::: ||||| |||||
Qy 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAla----- 94
      :::::||||| ||||| ||||| ::::: ||||| |||||
Db 372 TCTGACGATGAGTCCAAAGAAACATCTAAAGCTGCTCGCAGAGGTCAAGACCACTGAGAAA 431
      :::::||||| ||||| ||||| ::::: ||||| |||||
Qy 95 -----AspGlnThrAspAlaLeuGlnSerGluGluAlaGluAlaValValGln 109
      ||||| ||||| ||||| ||||| ||||| |||||
Db 432 ACAGCTGAAAACCCAGACGCTACAGATGACCAAGTAGCAAGACATCAGAGGGCGAGGAA 491
      ||||| ||||| ||||| ||||| ||||| |||||
Qy 110 SerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspVal 129
      ||||| ||||| ||||| ||||| ||||| |||||
Db 492 TTGGAT---TCAGCAGATGACAGCAAGACTCTACAGATGAGCCA---AGTCAAGAAACA 545
      :::::||||| ||||| ||||| ::::: ||||| |||||
Qy 130 LysLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsnValAsnThr 149
      :::::||||| ||||| ||||| ::::: ||||| |||||
Db 546 GCAGAGACTAAAGAAACAGAGTCAGCAGATAAAAGTAGTGAGGACAAAAGT-----TCAGCA 599
      :::::||||| ||||| ||||| ::::: ||||| |||||
Qy 150 LeuProThrThrGlyGluGluSerAsn 158
      ||||| ||||| ||||| ||||| |||||
Db 500 GCAGAGAAACTGATGAGGAAGCCTCA 626

RESULT 25
BJ000182          669 bp mRNA linear EST 05-DEC-2001
LOCUS             BJ000182 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA012D06 5',
DEFINITION        mRNA sequence.
ACCESSION         BJ000182
VERSION           BJ000182.1 GI:17355193
KEYWORDS          EST.
SOURCE            Japanese medaka.
ORGANISM          Oryzias latipes
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                  Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
                  Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE         1 (bases 1 to 669)
AUTHORS           Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE            Medaka EST Project in Takeda's lab
JOURNAL           Unpublished (2001)
COMMENT          Contact: Tadasu Shin-i
                  National Institute of Genetics
                  1111 Yata, Mishima, Shizuoka 411-8540, Japan
                  Tel: 81-559-81-6856
                  Fax: 81-559-81-6855
                  Email: tshini@genes.nig.ac.jp.
FEATURES          Location/Qualifiers
source            1..669
                  /organism="Oryzias latipes"
                  /strain="Hd-r"
                  /db_xref="taxon:8090"
                  /clone_lib="MF01SSA012D06"
                  /clone_lib="MF01SSA cDNA"
                  /sex="mixture of female and male"
                  /tissue_type="whole embryo"
                  /dev_stage="segmentation stage 20 - 25"
BASE COUNT        226 a 140 c 177 g 124 t
ORIGIN
Alignment Scores:
Pred. No.:        0.258 Length: 669
Score:            105.50 Matches: 34
Percent Similarity: 44.97% Conservative: 42
Best Local Similarity: 20.12% Mismatches: 74
Query Match:      13.58% Indels: 19
DB:               10 Gaps: 5

US-09-847-539A-6 (1-159) x BJ000182 (1-669)

Qy 2 AspSerProLeuGluGlnProArgileIleProAsnGlyGlyThrLeuThrAsnLeuLeu 21
      :::::||||| ||||| ||||| ::::: ||||| |||||

```

```

Db 160 GATGCTGGGACAAAAGAGCCAGCTGTTACAGAACGAGCTGAACAGCTGAACCTTCAGCT 219
Qy 22 GlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaLeuAspGluLeuLys 41
Db 220 GATTCATCTCACAGTGAATGCTATGATGACAACTCT-----GAGGACAACAGC 270
Qy 42 LysGlnAlaLeuGluAspLysGluAlaThrAlaLeuGluAlaLeuSerSerAspAla 61
Db 271 GATGCATCTATGATGAGCAAAATCTGAAGACAACGCTCAAGAGCTCCACTGAACATGATGAA 330
Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLeuValValLys 81
Db 331 GAAAAACCAAAATGACCAACTGACACCAACAGAGGAGGATTCACGATGAAGAAC 390
Qy 82 AlaAspAsnAlaLeuSerAspAlaLeuGluAlaLeuAla----- 94
Db 391 TCTCAGCATGAGTCCAAAGAAACATCTAAAGCTGCTGCAGAGTCAAGACCCTGAGAAA 450
Qy 95 -----AspGlnThrAspAlaLeuGlnSerGluGluAlaLeuValValLys 109
Db 451 ACAGCTGAAACCCAGAGCGCTACAGATGAACCAAGTGAAGAAACATCAGAGGCGAGGAA 510
Qy 110 SerAspAsnAlaLeuSerAspAlaLeuGluLysAlaLeuThrProLeuAlaLeuAspVal 129
Db 511 TTGGAT---TCAGCATGACGACCAAGCTCTACAGATGAGCCA---AGTCAAGAAACA 564
Qy 130 LysLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsnValAsnThr 149
Db 565 GCAGAGACTAAAGAACAGAGTCAAGAGATAAAAGTCAGGACCAAAAGT-----TCAGCA 618
Qy 150 LeuProThrThrGlyGluGluSerAsn 158
Db 619 GCAGAGAAACTGATGAAGAGCCTCA 645

RESULT 26
BJ011593
LOCUS
DEFINITION BJ011593 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA165C04 5',
mRNA sequence.
ACCESSION BJ011593
VERSION BJ011593
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM
Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 687)
Kohara,Y., Shin-I.,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
            Location/Qualifiers
     1..687
     /organism="Oryzias latipes"
     /strain="Hd-rR"
     /db_xref="taxon:8090"
     /clone="MF01SSA165C04"
     /clone_lib="MF01SSA cDNA"
     /sex="mixture of female and male"
     /tissue_type="whole embryo"
     /dev_stage="segmentation stage 20 - 25"

BASE COUNT 233 a 143 c 180 g 131 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 0.267 Length: 687
Score: 105.50 Matches: 34
Percent Similarity: 44.97% Conservative: 42
Best Local Similarity: 20.12% Mismatches: 74
Query Match: 13.58% Indels: 19
DB: 10 Gaps: 5

US-09-847-539A-6 (1-159) x BJ011593 (1-687)
Qy 2 AspSerProLeuGluGlnProArgIleileProAsnGlyGlyThrLeuThrAsnLeuLeu 21
Db 170 GATGCTGGGACAAAAGAGCCAGCTGTTACAGAACGAGCTGAACAGCTGAACCTTCAGCT 229
Qy 22 GlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaLeuAspGluLeuLys 41
Db 230 GATTCATCTCACAGTGAATGCTATGATGACAACTCT-----GAGGACAACAGC 280
Qy 42 LysGlnAlaLeuGluAspLysGluAlaThrAlaLeuGluAlaLeuSerSerAspAla 61
Db 281 GATGCATCTATGATGAGCAAAATCTGAAGACAACGCTCAAGAGCTCCACTGAACATGATGAA 340
Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLeuValValLys 81
Db 341 GAAAAACCAAAATGACCAACTGACACCAACAGAGGAGGATTCACGATGAAGAAC 400
Qy 82 AlaAspAsnAlaLeuSerAspAlaLeuGluAlaLeuAla----- 94
Db 401 TCTCAGCATGAGTCCAAAGAAACATCTAAAGCTGCTGCAGAGTCAAGACCCTGAGAAA 460
Qy 95 -----AspGlnThrAspAlaLeuGlnSerGluGluAlaLeuValValLys 109
Db 461 ACAGCTGAAACCCAGAGCGCTACAGATGAACCAAGTGAAGAAACATCAGAGGCGAGGAA 520
Qy 110 SerAspAsnAlaLeuSerAspAlaLeuGluLysAlaLeuThrProLeuAlaLeuAspVal 129
Db 521 TTGGAT---TCAGCATGACGACCAAGCTCTACAGATGAGCCA---AGTCAAGAAACA 574
Qy 130 LysLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsnValAsnThr 149
Db 575 GCAGAGACTAAAGAACAGAGTCAAGAGATAAAAGTCAGGACCAAAAGT-----TCAGCA 628
Qy 150 LeuProThrThrGlyGluGluSerAsn 158
Db 629 GCAGAGAAACTGATGAAGAGCCTCA 655

RESULT 27
BJ001283
LOCUS
DEFINITION BJ001283 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA005B12 5',
mRNA sequence.
ACCESSION BJ001283
VERSION BJ001283
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM
Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 710)
Kohara,Y., Shin-I.,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
            Location/Qualifiers
     1..710
     /organism="Oryzias latipes"
     source

```





```

Qy 112 Asn-----AlaAlaSerAspAlaTrpGluLysAlaAlaThrProileAlaLeuAsp 128
    ::::: ||||| ||| ::::: |||:::
Db 149 CCAGATTTCGTCAGTTTCGATGTCATCAGTTTCAGAGGACTGGCATTAGTAGTAA 90
    ::::: ||||| ||| ::::: |||:::

Qy 129 ValLysLysThrLysAspThrLysProValValLysLysGlu 142
    ::::: ||||| ||| ::::: |||:::
Db 89 TCGGAACCAAGAAATCGGCACCAGAGTAGAGAGCAGAA 48
    ::::: ||||| ||| ::::: |||:::

RESULT 29
BE405838
LOCUS BE405838 473 bp mRNA linear EST 21-JUL-2000
DEFINITION WHE0437_A10_A19ZS wheat etiolated seedling root cDNA library
Triticum aestivum cDNA clone WHE0437_A10_A19, mRNA sequence.
ACCESSION BE405838
VERSION BE405838.1 GI:9365306
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 473)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
Seaton,C.L. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat
genomes
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@nwp.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Strategene SK primer.
FEATURES
Location/Qualifiers
1..473
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE0437_A10_A19"
/clone_lib="Wheat etiolated seedling root cDNA library"
/tissue_type="Root"
/dev_stage="Five day old etiolated seedling"
/lab_host="E. coli SOLR"
/site="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: ECORI; Site_2: XhoI; Seeds were surface-sterilized
, germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered crystallization dishes. Roots were
harvested. The tissue, total RNA, and poly(A) RNA were
prepared, a cDNA library was made, and the cDNA clones
were in vivo excised to give phluescript phagemids in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
```

```

BASE COUNT 171 a 84 c 155 g 63 t
ORIGIN

Alignment Scores:
Pred. No.: 0.193 Length: 473
Score: 105.00 Matches: 40
Percent Similarity: 44.44% Conservative: 24
Best Local Similarity: 27.78% Mismatches: 60
Query Match: 13.51% Indels: 20
DB: 10 Gaps: 5
```

US-09-847-539A-6 (1-159) x BE405838 (1-473)

```

Qy 27 LysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGlu 46
    ||| ||| ||| ||| ::::: ||||| |||
Db 2 AAACCTGCCAAACGGGAAGGCCCTCAAGCTCTAAGAAAGGCANAAAGGGGAACAGGAG 61
    ||| ||| ||| ||| ::::: ||||| |||

Qy 47 AspLysGluAlaThrThrAlaIleGluAlaIleSerSerAspAlaLeuGluAlaLeuAla 66
    ||| ||| ||| ||| ::::: ||||| |||
Db 62 GATCAGAGGACGACGAGAGCTGAAAGTGGAGATCATCTGATGCTCAAGAACCAAGGC 121
    ||| ||| ||| ||| ::::: ||||| |||

Qy 67 AspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal----- 80
    ::::: ||| ::::: ||||| ||| |||
Db 122 ACTGAAGTGGAGATCAAGAGATGCTCAAGAAAGCCAAAGTCACTCAGCTGCACCTGCAGATC 181
    ||| ||| ||| ||| ::::: ||||| |||

Qy 81 ---LysAlaAspAlaAlaAlaSerAspAlaLeuGluAlaLeuAla----- 94
    ::::: ||||| ||| ||| |||
Db 182 AAGAAGCTGACANTGCTCAGGAAGAGAAAGGAAGGAGCGCTGGCGGCTCGGTGGA 241
    ||| ||| ||| ||| ::::: ||||| |||

Qy 95 -----AspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsn 112
    ::::: ||| ::::: ||||| ||| |||
Db 242 GAGAAGAAAGGCTCACACTCTTGCTCAGGAAGAGAGAGAAA-----CAGGCTCACACT 295
    ||| ||| ||| ||| ::::: ||||| |||

Qy 113 AlaAlaSerAspAlaTrpGluLysAlaAlaThrProileAlaLeuAspValLysThr 132
    ||| ||| ||| ||| ::::: ||||| |||
Db 296 GTTGCTCAGGAAGAG---AAGAAGAGGCGCCACCTGTTGCTCAAGAGAGAGAAAGAG 352
    ||| ||| ||| ||| ::::: ||||| |||

Qy 133 LysAspThrLysProValValLysLysGluGluArgGlnAsnValAsnThrLysProThr 152
    ||| ||| ||| ||| ::::: ||||| |||
Db 353 GCGGACACT-----CTTGCTCAAGAGAGAGAAAGAGGCTGCACACTGCTGCTGAA 403
    ||| ||| ||| ||| ::::: ||||| |||

Qy 153 ThrGlyGluGlu 156
    ::::: |||
Db 404 GAAGAGAGAGAA 415
    ::::: |||

RESULT 30
AJ395963
LOCUS AJ395963 755 bp mRNA linear EST 25-JAN-2001
DEFINITION AJ395963 dkfz426 Gallus gallus cDNA clone 24ml7r1, mRNA sequence.
ACCESSION AJ395963
VERSION AJ395963.1 GI:7127417
KEYWORDS EST.
SOURCE Chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 755)
AUTHORS Abdrakhmanov,I., Lodygin,D., Gerroth,P., Arakawa,H., Law,A., Plachy
,J., Korn,B. and Buerstedde,J.M.
TITLE A large database of chicken bursal ESTs as a resource for the
analysis of vertebrate gene function
JOURNAL Genome Res. 10 (12), 2062-2069 (2000)
MEDLINE 20568495
COMMENT Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
Location/Qualifiers
1..755
/organism="Gallus gallus"
/strain="CB"
/db_xref="taxon:9031"
/clone="24ml7r1"
/clone_lib="dkfz426"
/tissue_type="Bursa of Fabricius"
BASE COUNT 315 a 124 c 185 g 131 t
ORIGIN

Alignment Scores:
Pred. No.: 0.335 Length: 755
Score: 105.00 Matches: 34
Percent Similarity: 48.15% Conservative: 31
Best Local Similarity: 25.19% Mismatches: 56
```

```
Query Match: 13.51% Indels: 12
DB: 9 Gaps: 4

US-09-847-539A-6 (1-159) x AJ395963 (1-755)

QY 17 LeuThrAsnLeuLeuGlyAsnAlaProGluLysLeuAla-----LeuArgAsnGluGlu 34
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 TTGGAGATAAATGAAATGAAATGCAAGACGACGCGTGGGAAGGAAATAAATAATGCCAG 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 35 ArgAlaIleAspGluLysGluAlaIleGluAspLysGluAlaThrThrAlaIle 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 CAGAACTAAATAGCCCAAGAAAGACGAGATTCAAGCAGAAAAATG----- 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 55 GluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 ---AAAGAAAGCAGCAAGAGTGAAGCATTAGTTCTGGAACCTTGAACAGCTGAACAA 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 75 GluGluAlaAlaValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAla 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 GAACAAGCCTCATATAGCAACAGAGTGAGCGACGACAAACAAAGCAATCGCATCTTTAAAG 312
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 95 AspGlnThrAspAlaLeuGlnSerGluGluAlaGluValGlnSerAspAsnAlaAla 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 GACCAAGTCAGTCCTTG-----GAAGCTGAGCGAGTGAAGACAGGGAATCTCTA 363
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 115 SerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysThrLysAsp 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 AAGAAGCAGAGATGAGCTGCTCCATGAAAGGGATTATGCGCAGAACGAACTAAAGAT 423
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 135 ThrLys-----ProValValLysLysGluGluArgGlnAsn 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 ATTAAGCCAAATCTGCAAGATAGAGAAATACAGAGAGCAAAAT 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 31
BM300555
LOCUS
DEFINITION
MCA053H02.21480 Ice plant Lambda Uni-zap XR expression library, 0
hours NaCl treatment prescreened for removal of highly abundant
transcripts Mesembryanthemum crystallinum cDNA clone MCA053H02 5,
mRNA sequence.
BM300555
VERSION 530 bp mRNA linear EST 22-JAN-2002
KEYWORDS MCA053H02.21480 Ice plant Lambda Uni-zap XR expression library, 0
SOURCE hours NaCl treatment prescreened for removal of highly abundant
transcripts Mesembryanthemum crystallinum cDNA clone MCA053H02 5,
ORGANISM mRNA sequence.
Mesembryanthemum crystallinum
common ice plant.
Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
1 (bases 1 to 530)
Cushman, J.C.
An expressed sequence tag database for the common ice plant,
Mesembryanthemum crystallinum
Unpublished (1997)
Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR PRIMERS
FORWARD: T3 20mer
BACKWARD: T7 21mer
Plate: 053 row: H column: 02
Seq primer: T3 20mer
High quality sequence stop: 530.
Location/Qualifiers
1..530
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone="MCA053H02"
, 0 hours NaCl treatment prescreened for removal of highly

FEATURES
source
```

```
abundant transcripts"
/dev_stage="six-week-old"
/note="Vector: Lambda Uni-zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 186 a 99 c 173 g 72 t
ORIGIN
Alignment Scores:
Pred. No.: 0.28 Length: 530
Score: 104.00 Matches: 44
Percent Similarity: 40.13% Conservatives: 17
Best Local Similarity: 28.95% Mismatches: 49
Query Match: 13.38% Indels: 42
DB: 10 Gaps: 7

US-09-847-539A-6 (1-159) x BM300555 (1-530)
QY 30 LeuArgAsnGluGluArgAlaIleAspGluLysGlnAlaIleGluAspLysGlu 49
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9 TTGACGAACGAGGAGGGGCG-----GCGCGCGCGCGGAGGAGGAAGCCA 53
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 50 AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54 GCA-----GCGGTGGAAGTTGCGCGCGGGAAGGTGGCGAAGTGTGCGGAAAGTAGA 107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 AspAla-----LeuGlnSerGluGluAlaAlaValLysAlaAspAla 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 GATATTGCGGTGCTGACTTTAAAAAGGAGGAAGAAAGAACGCCGAGCTGAGGAGGCC 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168 GCAGAG-----CCACCGGAGGAGAG 188
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 GCTGCAGAGCCAGCCAAAGAAAGCCGAAAGCCAGCAGAGGAAAGGCTGTGAGCCAGCC 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 126 AlaLeuAspValLysThrLysAsp-----Thr 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 ACCAAGAAGCTGAGACCAAGAAAGAAAGAGTCCCAAGCCAGCTGAAGCTGAACCGGCT 308
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 136 LysProValVal-----LysLysGluGluArgGlnAsnValAsnThrLeuProThr 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 AAGCCAGTACCCCGCCGAGCAAGAGAGGAGGAGCAAGCCAGCTTGGCGAGCTCCAGCA 368
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 153 ThrGly-----GluGluSerAsnPro 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 GCGGCTGCTGAGAAGAAGGTTGAGGAGAGCAAAACCC 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 32
BJ006142
LOCUS
DEFINITION
BJ006142 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA089B12 5',
mRNA sequence.
ACCESSION BJ006142
VERSION BJ006142.1 GI:17359859
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 616)
Kohara, Y., Shin-i, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.
Medaka EST project in Takeda's lab
Unpublished (2001)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
```



ORGANISM Entamoeba histolytica  
Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 938)  
AUTHORS Loftus B., Van Aken, S. and Fraser C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
JOURNAL HMI:IMSS sheared DNA library  
COMMENT Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjl@fetusitigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library  
Seq primer: M13-Reverse  
Class: shotgun  
High quality sequence start: 34  
High quality sequence stop: 809.

FEATURES  
source  
1..938  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: PHOS1; Site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell, Oxford University Press, 1999)."

BASE COUNT 132 a 201 c 107 g 498 t

Alignment Scores:  
Pred. No.: 0.697 Length: 938  
Score: 103.00 Matches: 40  
Percent Similarity: 49.62% Conservative: 26  
Best Local Similarity: 30.08% Mismatches: 59  
Query Match: 13.26% Indels: 8  
DB: 12 Gaps: 4

US-09-847-539A-6 (1-159) x AZ676392 (1-938)

Qy 16 ThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArg 35  
Db 829 AGTCTAAGAAGAAGCTGAAGCAAAAGCCAAAGAAAGAGCT-----GAAGCAAAA 779

Qy 36 AlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGlu 55  
Db 778 GCTAAGAAGAAGCTGAAGCAAAAGCTTAAAGAAGAAGCTGAAGCAAAAGCTAAGAAGAA 719

Qy 56 AlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 75  
Db 718 GCTGAACCAAAAGCTTAAAGAAGAAGCTGAAGCAAAAGCCAAAGAA-----GAAGCTGAA 665

Qy 76 GluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAla--- 94  
Db 664 GCAAAAGCTTAAAGAAGAAGCTGAAGCAAAAGCCAAAGAAAGCTGAAGCAAAAGCTAAA 605

Qy 95 AspGlnThrAspAlaLeuGlnSerGluGluAlaValValGlnSerAspAsnAlaAla 114  
Db 604 GAAGAAGCTGAAGCAAAAGCTTAAAGAAGAAGCTGAAGCAAAAGCTAAGAAGCTGAAG 545

Qy 115 SerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThrLysAsp 134

Db 544 GCAAAAGCTTAAAGAAGAAGCTGAAGCAAAAGCTTAAAGAAGCTTAA-----GAAGAA 491

Qy 135 ThrLysProValValLysLysGluGluArgGlnAsnVal 147  
Db 490 AATAAAGAAGAAGCTTAAAGAAGAAGCTTAAAGAAGAAGCTT 452

RESULT 35  
BG930118 537 bp mRNA linear EST 31-DEC-2001  
LOCUS eusHEST0539 EctH11 Eimeria tenella cDNA clone etsHK077 5', mRNA  
DEFINITION sequence.  
ACCESSION BG930118  
VERSION BG930118.1 GI:18003508  
KEYWORDS EST.  
SOURCE Eimeria tenella.  
ORGANISM Eimeria tenella.  
REFERENCE 1 (bases 1 to 537)  
AUTHORS Ng, S.T., Jangi, M.S., Shirley, M.W., Tonley, F.M. and Wan, K.L.  
TITLE Comparative EST analyses provide insights into gene expression in  
two asexual developmental stages of Eimeria tenella  
JOURNAL Unpublished (2001)  
COMMENT Contact: Wan KL  
Centre for Gene Analysis and Technology  
Universiti Kebangsaan Malaysia  
43600 UKM Bangi, Selangor DE, Malaysia  
Tel: 6 03 8292997  
Fax: 6 03 8293249  
Email: kwan@krisc.cc.ukm.my  
PCR Primers  
FORWARD: T3  
BACKWARD: T7  
Seq primer: SK.

FEATURES  
source  
1..537  
/organism="Eimeria tenella"  
/strain="Houghton"  
/db\_xref="taxon:5802"  
/clone\_lib="etsHK077"  
/dev\_stage="Sporozoite"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: Lambda ZAPII; Site\_1: EcoRI; Site\_2: XhoI;  
Sporozoites were excysted in vitro from E. tenella H  
oocysts, purified by column chromatography and mRNA  
extracted using a FASTTRACK kit (Invitrogen). cDNA was  
synthesised and a Uni-ZAP XR library was constructed  
using cDNA synthesis kit, ZAP-cDNA synthesis kit and  
ZAP-cDNA Gigapack III Gold cloning kit (Stratagene). The  
library was amplified once through E. coli XLI-Blue  
MRF".

BASE COUNT 175 a 140 c 191 g 29 t 2 others

ORIGIN

Alignment Scores:  
Pred. No.: 0.405 Length: 537  
Score: 102.50 Matches: 42  
Percent Similarity: 42.42% Conservative: 28  
Best Local Similarity: 25.45% Mismatches: 61  
Query Match: 13.19% Indels: 34  
DB: 10 Gaps: 5

US-09-847-539A-6 (1-159) x BG930118 (1-537)

Qy 29 AlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLys 48  
Db 8 GCACACACACACAGGAGGAGCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 67

Qy 49 GluAlaThrThrAla-----IleGluAlaAlaSerSerAspAlaLeuGlu----- 63  
Db 68 GAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 127

QY 64 -----AlaLeuAlaAspGln 68  
 Db 128 CGCTACGGCTCTGGCGTGAAGACGAGCTGAAGAGAGAAAGCGGCACGAGACTCT 187  
 QY 69 ThrAsp-----AlaLeuGlnSerGluAlaAlaValValLysAlaAspAsn 84  
 Db 188 GGAGAGAGAGACGACGACGAGCGAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAG 247  
 QY 85 AlaAlaSerAspAlaLeuAlaLeuAlaAspGln-----ThrAspAlaLeu 100  
 Db 248 CCGGACGAGAGTGTTCGGAAGACGAGCAGAGCTCAGGGCGCGGCACGACGACAGA 307  
 QY 101 GlnSerGluAlaAlaGluValValGlnSerAspAsn-----AlaAlaSer 115  
 Db 308 GAGTCGAGAGAGAGACGACGACGAGAGACTCTGACGACGAGCGACGACGAGAA 367  
 QY 116 AspAlaTrpGluLysAlaAlaThrProLysAlaLeuAspValLysThrLysAspTh 135  
 Db 368 GACTCTGAGGAGAGACGACGACGAGAGAGTCCCGCGGATGCGACGACGAGCG 427  
 QY 135 rLysProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGl 155  
 Db 428 GAGGACCGGACGACGACGACGAGAGAGTCCCGCGGATGCGACGAGAGAGAGTCC 487  
 QY 155 uGluSerAsnPro 159  
 Db 488 GCAGCAGACTCCA 500

RESULT 36  
 A0648620  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Other\_GSSs: RPCI93-EcoRI-6D15.TJ  
 Contact: Najib M. El-Sayed  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: nelsayed@tigr.org

Clones and high density filters may be purchased from BACPAC  
 Resources (<http://bacpac.med.buffalo.edu>). BAC end sequences search  
 page: <http://www.tigr.org/tldb/mdb/tbdb/>.  
 Seq primer: 17  
 Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
 1..628  
 /organism="Trypanosoma brucei"  
 /strain="TREU927/4 GUTat 10.1"  
 /db\_xref="taxon:5691"  
 /clone="RPCI93-EcoRI-6D15"  
 /note="Vector: pBACE3.6; Site.1: Eco RI; Site.2: Eco RI;  
 Constructed for The Institute for Genomic Research by  
 Bohui Zhao in Pieter de Jong's laboratory (Roswell Park  
 Cancer Institute, Buffalo, NY). Briefly, Trypanosoma  
 brucei TREU927/4 GUTat 10.1 agarose embedded DNA was

partially digested with a combination of Eco RI and Eco RI  
 methylase (Rpci93-EcoRI segment) or Dpn II (Rpci93-DpnII  
 segment). High molecular weight fragments were ligated in  
 pBACE3.6 vector digested with Eco RI or Bam HI,  
 respectively. The average insert size is 141 Kb. Total  
 coverage (both segments): > 90 X the haploid  
 non-minichromosomal genome."

BASE COUNT 218 a 127 c 168 g 114 t 1 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.488 Length: 628  
 Score: 102.50 Matches: 47  
 Percent Similarity: 44.44% Conservative: 33  
 Best Local Similarity: 26.11% Mismatches: 61  
 Query Match: 13.19% Indels: 39  
 DB: 12 Gaps: 8

US-09-847-539A-6 (1-159) x A0648620 (1-628)

QY 1 ValAspSerProGluGlnProArgGluLeuProAsnGlyGlyThrLeuThrAsnLeu 20  
 Db 39 GTCTGAGATAGCTCGCTGAACCCCGT-----CGTCTTCTGAAGAA 80  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArg----- 35  
 Db 81 ATAGACAACCGTGTGCAAAATTTGAAGGTGACTGTGGATGAAATAATGCGGAGAAGTCA 140  
 QY 36 ---AlaLeuAspGluLysLysGlnAlaLeuGluAspLysGlnAlaThrThrAlaLeu 54  
 Db 141 GCTGCTGAGGATGAGCTTCAACACGGAAGAAGAACTGCTCAACACGACGAGGAGCGT 200  
 QY 55 GluAlaAlaSerSer-----AspAlaLeuGluAlaLeu 65  
 Db 201 GAGGACGCGCTACAAAAGACAGACGTTGCTCAGAGAGGAGTGGTGGTGGTGGTGGG 260  
 QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 85  
 Db 261 GCGGATGAGCAGCTCGAAGTCAAGGAAGAAAGAAAGAAACAAATAA-----GATGCT 314  
 QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 102  
 Db 315 GCAAGCGCATGCAAAAGGCTCAACTACAAACAGGAGCGGATCGAGAGGCTGCCATC 374  
 QY 103 -----GluGluAlaGluValValGluSerAsp---Asn 112  
 Db 375 CATCAGCAGATCTCAATGAAGCGCGAGAACAGTAGAATACTTCAGAAAGAGTATGAC 434  
 QY 113 AlaAlaSerAspAlaTrpGluLysAlaAlaThrProLeuAlaLeuAspVal---LysLys 131  
 Db 435 GTCCGACACGAGAGGTTTGTAGCATCTCACACCACITTCGCCAGGAGCTAGATAAGAAG 494  
 QY 132 ThrLysAspThrLysProValValLysLys-----GluGluArgGlnAsnValAsnThr 149  
 Db 495 AAGGCTGATTTTGCCTCATATGAAAAGCTCTGCTGGAAGCAAGAGCAACTGATGACG 554

RESULT 37  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

CNS07270 790 bp DNA linear GSS 07-JUL-2001  
 clone BA0AB008E01 of library BA0AB from strain CLIB 210 of  
 Kluyveromyces lactis, genomic survey sequence.

AL425866  
 AL425866.1 GI:12209060  
 GSS.  
 Kluyveromyces lactis.  
 Kluyveromyces lactis

REFERENCE  
 AUTHORS

1 (bases 1 to 790)  
 Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 Saccharomycetaceae; Kluyveromyces.  
 Bolotin-Fukuhara, M., Toffano-Nioche, C., Artiguenave, F.,  
 Duchateau-Nguyen, G., Lemaire, M., Marmelisse, R., Montrocher, R.,  
 Robert, C., Terrier, M., Winkler, P., and Wesolowski-Louvel, M.  
 Genomic exploration of the hemiascomycetous yeasts: 11.

LOCUS	BG440614	727 bp	mRNA	linear	EST 15-MAR-2000
DEFINITION	GA_Ea0009A22f Gossypium arboreum 7-10 dpa fiber library Gossypium				
ACCESSION	arboresm cdna clone GA_Ea0009A22f, mRNA sequence.				
VERSION	BG440614				
KEYWORDS	BG440614.1	GI:13350265			
SOURCE	EST.				
ORGANISM	Gossypium arboreum.				
	Gossypium arboreum				
	Eukaryota; Viridiplantae;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.				
REFERENCE	1 (bases 1 to 727)				
AUTHORS	Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry				
TITLE	D., Wood,T.C., Leslie,A. and Wilkins,T.A.				
JOURNAL	An integrated analysis of the genetics, development, and evolution				
COMMENT	of the cotton fiber				
	Unpublished (2000)				
	Contact: Wing RA				
	Clemson University Genomics Institute				
	Clemson University				
	100 Jordan Hall, Clemson, SC 29634, USA				
	Tel: 864 656 7288				
	Fax: 864 656 4293				
	Email: rwing@clemson.edu				
	Seq primer: TAATACGACTCACTATAGG				
	High quality sequence stop: 504.				
FEATURES	Location/Qualifiers				
- source	1..727				
	/organism="Gossypium arboreum"				
	/strain="AKA"				
	/cultivar="8400"				
	/db_xref="taxon:29729"				
	/clone_lib="GA_Ea0009A22f"				
	/clone="GA_Ea0009A22f"				
	/tissue_type="Fibers isolated from bolls harvested 7-10				
	dpa"				
	/lab_host="E. coli"				
	/note="Vector: pBK-CMV; Site_1: ECORI; Site_2: XhoI"				
BASE COUNT	233 a	138 c	201 g	153 t	2 others
ORIGIN					
Alignment Scores:					
Pred. No.:	0.653	Length:	727		
Score:	102.00	Matches:	41		
Percent Similarity:	43.48%	Conservative:	29		
Best Local Similarity:	25.47%	Mismatches:	51		
Query Match:	13.13%	Indels:	40		
DB:	10	Gaps:	6		
 US-09-847-539A-6 (1-159) x BG440614 (1-727)					
Qy	9	ArgilleleproasnGlyGlyThrLeuThrasnLeuLeuGlyasnalaproGluLysLeu	28		
Db	23	CAAFATTTTCCCACTTCCTTTTGCCCTCAAATAATTCCTGGGATTCCTATGCCTACTGCT	82		
Qy	29	AlaLeuArgAsnGlGuAlarGalaleaspGluLeuLys---	LysGlnAlarleGluasp	47	
Db	83	GAGGTGCCATCAGTTCAAACCTGCACCTTCATGAGGAGAAAACCTGAGATCATGATGAGGAT	142		
Qy	48	LysGluAlaThrThrAlatleGluAlaLaSerSeraspAlaLeuGluAlaLeuAlaasp	67		
Db	143	CAAGAGCTACAACAGATGAAGTAGTTGCCGAAGCTCCTCGACGGAAGCAGCTGCCAA	202		
Qy	68	GlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAlaAspAsnAlaSer	87		
Db	203	GAGCCAAAG-----GAAGCGAGAGCTCGCGCGTGGCAGTCGAGGAACTGAAGCCTCG	256		
Qy	88	AspAlaLeuGluAlaLeuAlaaspGlnThrAspAlaLeuGlnSerGluGluAlaGluval	107		
Db	257	-----GTTGAAGTTGAAACCAGAGAGGTG	280		
Qy	108	valcInserAspsnAlaAlaserasp-----AlatrplCgluys	120		
	:		: :: :		

```
Db 281 GTAGAAGAACCAAGGCTGTTACTAGTAGGAGCCAGTAGTAGAGGAGACTCCTTAAGGAACA 340
Qy 121 AlaAlaThrProIleAlaLeuAspVallys-----LysThrLys 133
Db 341 GTGCCAGAACCAAGTTCGCCGAGGAGATCAAGGAACCACTGAACANACCGTTGAACCAAA 400
Qy 134 AspThr-----LysProValVallysGlu 142
Db 401 GAAACTACAGAGTCTAGTAGGAGCAGCAGAGAGAGCACCACCAAGAGAGAGTTGAGAAGAA 460
Qy 143 Glu 143
Db 461 GAA 463

RESULT 39
CNS06MOG/c 1076 bp DNA linear GSS 17-JUN-2001
LOCUS T3 end of clone AU0AA006A09 of library AU0AA from strain CBS 3082
DEFINITION of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION AL405734
VERSION AL405734.1 GI:12168821
KEYWORDS GSS.
SOURCE Saccharomyces kluyveri.
ORGANISM Saccharomyces kluyveri.
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 1076)
Neueglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
Gaillardin,C. and Casaregola,S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 9.
JOURNAL Saccharomyces kluyveri
MEDLINE FEBS Lett. 487 (1), 56-60 (2000)
REFERENCE 20584719
AUTHORS 2 (bases 1 to 1076)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Botolin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Liorente,B.,
Maupertuy,A., Neueglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
REFERENCE 3 (bases 1 to 1076)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, Cp 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
location/Qualifiers
1..1076
/organism="Saccharomyces kluyveri"
/strain="CBS 3082"
/db_xref="taxon:4934"
/clone="AU0AA006A09"
/clone_lib="AU0AA"
/notes="end : T3"
BASE COUNT 221 a 290 c 180 g 383 t 2 others
ORIGIN
Alignment Scores: 1.04 Length: 1076
Pred. No.: . . . .
```

```
Score: 102.00 Matches: 35
Percent Similarity: 50.81% Conservative: 28
Best Local Similarity: 28.23% Mismatches: 49
Query Match: 13.13% Indels: 12
DB: 12 Gaps: 4

US-09-847-539A-6 (1-159) x CNS06MOG (1-1076)
Qy 23 AsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLys 42
Db 832 GATGCTGAAGAAACCGAAGAGCGGAGACCAAGAGAGCTTCATTGGCAGCAGATTCTAGG 773
Qy 43 GlnAlaIle---GluAspLysGluAlaThrAlaIleGluAlaAala-----Ser 58
Db 772 GAAGCACTAGCGGAAGACGAGAGAGCTGAGGCTACCGAAGCGCAGCTGGAAGAGAT 713
Qy 59 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 78
Db 712 TGTGAGGCTGCTACCGCGCTGACAGAGATGCGGACATTTCTAGAAGCTACTGAAGCTTCA 653
Qy 79 ValVal-----LysAlaAspAsnAlaAlaIleSerAspAlaLeuGluAlaLeuAla 94
Db 652 CTGTTAGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 593
Qy 95 AspGlnThrAspAlaLeuGlnSerGluGluAla-----GluValValGlnSer 110
Db 592 GCGGAAGCAGAAAGCAACAGCAAGCAACAGCAAGCAAGCAAGCAAGCAAGCAAGCG 533
Qy 111 AspAsnAlaIleSerAspAlaThrGluLysAlaAlaThrProIleAlaLeuAspVallys 130
Db 532 GATGACGCGCTAGCAGAGCAGCGGAATCAGAAAGCAACAGCAAGCAGATCGCGCTAGCA 473
Qy 131 LysThrLysAsp 134
Db 472 CTGCTCAAAAGAT 461

RESULT 40
BH161951/c
LOCUS BH161951
DEFINITION ENTFRB12TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION BH161951
VERSION BH161951.1 GI:15735389
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 868)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 13
High quality sequence stop: 861.
location/Qualifiers
1..868
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: PHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
```



Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999).

BASE COUNT 97 a 200 c 99 g 472 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.907 Length: 868  
Score: 101.50 Matches: 38  
Percent Similarity: 50.82% Conservative: 24  
Best Local Similarity: 31.15% Mismatches: 51  
Query Match: 13.06% Indels: 9  
DB: 12 Gaps: 4

US-09-847-539A-6 (1-159) x BH161951 (1-868)

```
Qy 33 GluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThr 52
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 610 GAAGCAAAAGCCAAAGAGAGCTGAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCT 551

Qy 53 AlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAla-----AspGlnThr 69
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 550 AAAGCAAGAGCTGAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCTAAAGAGAGAGCT 491

Qy 70 AspAlaLeuGlnSerGluGluAla-----AlaValValLysAlaAspAsnAlaAla 86
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 490 GAAGCAAAAGCCAAAGAGAGAGCTGAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCC 431

Qy 87 SerAspAlaLeuGluAlaLeuAla---AspGlnThrAspAlaLeuGlnSerGluGluAla 105
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 430 AAAGCAAGAGCTGAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCTAAAGAGAGAGCT 371

Qy 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 370 GAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCT 311

Qy 126 AlaLeuAspValLysThrLysAspThrLysProValValLysLysGluGluArgGln 145
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 310 AAAGAGAGAGTTAA-----GAAGAAATAAAGAGAGTTAAAGAGAGTTAAAGAA 257

Qy 146 AsnVal 147
   |||
Db 256 GAAGTT 251
```

Search completed: October 13, 2002, 06:10:13  
Job time : 1886.39 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:46:02 ; Search time 68,8717 Seconds  
(without alignments)  
399.383 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQRIIPNGTGLTNL.....KKEERQNVNLTPTTGERSNP 159

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	777	100.0	217	16 Q9X5C4	Q9X5C4 streptococc
2	760	97.8	156	2 Q9S6G4	Q9S6G4 streptococc
3	754	97.0	156	2 Q9S6G3	Q9S6G3 streptococc
4	694	89.3	268	2 Q9X5C6	Q9X5C6 streptococc
5	692	89.1	239	2 Q9X5C5	Q9X5C5 streptococc
6	151.5	19.5	439	2 Q9S105	Q9S105 streptococc
7	142	18.3	103	2 Q9S180	Q9S180 streptococc
8	138.5	17.8	323	2 Q9S390	Q9S390 streptococc
9	122.5	15.8	1110	13 Q9X4M4	Q9X4M4 caenorhabdi
10	117	15.1	7659	5 Q9X4M4	Q9X4M4 caenorhabdi
11	116	14.9	166	10 Q9SCK5	Q9SCK5 arabidopsis
12	114.5	14.7	413	2 Q9S974	Q9S974 streptococc
13	114.5	14.7	850	2 Q9S5B1	Q9S5B1 salmonella
14	112.5	14.5	425	2 Q94744	Q94744 streptococc
15	111	14.3	388	2 Q9AMM4	Q9AMM4 streptococc
16	111	14.3	415	2 Q94859	Q94859 streptococc

17 111 14.3 619 5 Q9S5L8  
18 111 14.3 2055 2 Q9S472  
19 110 14.2 890 2 Q9L891  
20 109 14.0 316 10 Q9LK53  
21 109 14.0 425 2 Q94860  
22 108 13.9 159 10 Q23881  
23 108 13.9 405 2 Q94862  
24 108 13.9 415 2 Q9S5L8  
25 108 13.9 425 2 Q9S808  
26 108 13.9 655 2 Q9FIF5  
27 108 13.9 679 2 Q9Z4H8  
28 108 13.9 1096 5 Q94478  
29 108 13.9 1096 5 Q9VFAL  
30 107 13.8 384 2 Q9S810  
31 107 13.8 388 2 Q93474  
32 106.5 13.7 437 2 Q9EY8  
33 105.5 13.6 314 2 P71213  
34 105.5 13.6 389 2 Q9ZIX5  
35 105 13.5 387 2 Q94842  
36 105 13.5 524 2 Q9K2J9  
37 105 13.5 664 2 Q93975  
38 104.5 13.4 843 2 Q47802  
39 104.5 13.4 924 5 Q9BHC4  
40 104.5 13.4 1822 2 Q07290  
41 104 13.4 576 2 Q9KJJ3  
42 104 13.4 891 2 Q04111  
43 104 13.4 1795 2 Q9LCJ9  
44 104 13.4 2478 2 Q9RL69  
45 104 13.4 2481 16 Q99QR6

#### ALIGNMENTS

#### RESULT 1

Q9X5C4 PRELIMINARY; PRT; 217 AA.  
ID Q9X5C4  
AC Q9X5C4  
DT 01-NOV-1999 (TREMREL. 12, Created)  
DT 01-NOV-1999 (TREMREL. 12, Last sequence update)  
DE 01-DEC-2001 (TREMREL. 19, Last annotation update)  
DE GRAB PRECURSOR (PROTEIN GRAB) (PROTEIN G-RELATED ALPHA 2M-BINDING DE PROTEIN).  
GN GRAB OR SPV1357.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC700294;  
RA Rasmussen M., Muller H.P., Bjorck L.;  
RT "protein GRAB of Streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin."  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvarov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
DR EMBL: AF124399; AAD26338.1;  
DR EMBL: AE006573; AAK34185.1;  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
DR PRINTS: PR00015; GPOSANCHOR.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Complete proteome; Signal; Transmembrane.  
FT SIGNAL 1 33 POTENTIAL.

Q9S5L8 drosophila  
Q9S472 abiotrophia  
Q01891 enterococcu  
Q9LK53 arabidopsis  
Q54860 streptococc  
Q23881 fagopyrum e  
Q54862 streptococc  
P9S813 streptococc  
P9S808 streptococc  
Q9FIF5 enterococcu  
Q9Z4H8 enterococcu  
Q24478 drosophila  
Q9VFAL drosophila  
P9S810 streptococc  
Q53474 streptococc  
Q9EY8 escherichia  
P71213 escherichia  
Q9ZIX5 streptococc  
Q54842 streptococc  
Q9K2J9 enterococcu  
Q53975 streptococc  
Q47802 enterococcu  
Q9BHC4 leishmania  
Q07290 streptococc  
Q9KJJ3 enterococcu  
Q04111 enterococcu  
Q9LCJ9 staphylococ  
Q9RL69 staphylococ  
Q99QR6 staphylococ

```

FT CHAIN 34 217 GRAB.
SQ SEQUENCE 217 AA; 22836 MW; 79A8AC4FF5F3FA06 CRC64;

Query Match 100.0%; Score 777; DB 16; Length 217;
Best Local Similarity 100.0%; Pred. No. 2.3e-43;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60
Db 34 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 93

Qy 61 ALEALADQTDALQSEEAHVKNADNAADALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 94 ALEALADQTDALQSEEAHVKNADNAADALEALADQTDALQSEEAHVQSDNAASDAWEK 153

Qy 121 AATPIALDVKKTKDKTPVVKKEERQNVNLTPTTGEENP 159
Db 154 AATPIALDVKKTKDKTPVVKKEERQNVNLTPTTGEENP 192

RESULT 2
Q9S6G4 ID Q9S6G4 PRELIMINARY; PRT; 156 AA.
AC Q9S6G4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124401; AAD26340.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1 156
FT NON_TER 156 156
SQ SEQUENCE 156 AA; 16575 MW; 0C20967E2FB0D866 CRC64;

Query Match 97.8%; Score 760; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 2e-42;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60

Qy 61 ALEALADQTDALQSEEAHVKNADNAADALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 61 ALEALADQTDALQSEEAHVKNADNAADALEALADQTDALQSEEAHVQSDNAASDAWEK 120

Qy 121 AATPIALDVKKTKDKTPVVKKEERQNVNLTPTTGEENP 156
Db 121 AATPIALDVKKTKDKTPVVKKEERQNVNLTPTTGEENP 156

RESULT 3
Q9S6G3 ID Q9S6G3 PRELIMINARY; PRT; 156 AA.
AC Q9S6G3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.

```

```

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124401; AAD26339.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1 156
FT NON_TER 156 156
SQ SEQUENCE 156 AA; 16587 MW; 19FAA9800599D866 CRC64;

Query Match 97.0%; Score 754; DB 2; Length 156;
Best Local Similarity 99.4%; Pred. No. 5e-42;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60

Qy 61 ALEALADQTDALQSEEAHVKNADNAADALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 61 ALEALADQTDALQSEEAHVKNADNAADALEALADQTDALQSEEAHVQSDNAASDAWEK 120

Qy 121 AATPIALDVKKTKDKTPVVKKEERQNVNLTPTTGEENP 156
Db 121 AATPIALDVKKTKDKTPVVKKEERQNVNLTPTTGEENP 156

RESULT 4
Q9X5C6 ID Q9X5C6 PRELIMINARY; PRT; 268 AA.
AC Q9X5C6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124403; AAD26342.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1 268
FT NON_TER 268 268
SQ SEQUENCE 268 AA; 1C59239260CDC7E7 CRC64;

Query Match 89.3%; Score 694; DB 2; Length 268;
Best Local Similarity 58.2%; Pred. No. 7.1e-38;
Matches 156; Conservative 0; Mismatches 0; Indels 112; Gaps 1;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAAS-- 58
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60

Qy 59 -----
Db 61 ALEALADQADALQSEEAHVQSDNAADALEALADQADALQSEEAHVQSDNAAGDALEA 120

```

```
Qy 59 -----SDALEALADQ 68
Db 121 LADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDALEALADQ 180
Qy 69 TDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDALEALADQ 128
Db 181 TDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDALEALADQ 240
Qy 129 VKTKDKTPVVKKEERONVNTLPTTGE 156
Db 241 VKTKDKTPVVKKEERONVNTLPTTGE 268

RESULT 5
Q9X5C5
ID Q9X5C5 PRELIMINARY; PRT; 239 AA.
AC Q9X5C5;
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT)
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1314;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=KTL9;
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
RT bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124402; AAD26341.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 239
FT SEQUENCE 239 AA; 24846 MW; 61AC4F6F863AF0F5 CRC64;
SQ SEQUENCE 239 AA; 24846 MW; 61AC4F6F863AF0F5 CRC64;

Query Match 89.1%; Score 692; DB 2; Length 239;
Best Local Similarity 64.0%; Pred. No. 8.4e-38;
Matches 153; Conservative 0; Mismatches 2; Indels 84; Gaps 1;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPKALNEERAIDELKKQAEIDKEATTAEEAASD 60
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPKALNEERAIDELKKQAEIDKEATTAEEAASD 60
Qy 61 A-----
Db 61 ALEALADQADALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDITLEA 120
Qy 62 -----LEALADQTDALQSEAAVVKADNAASDALEALADQ 96
Db 121 LADQTDALQSEAAVVKADNAASDITLEALADQTDALQSEAAVVKADNAASDITLEALADQ 180
Qy 97 TDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDITLEALADQ 155
Db 181 TDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDITLEALADQ 239

RESULT 6
Q55105
ID Q55105 PRELIMINARY; PRT; 439 AA.
AC Q55105;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE MULTIPLE LIGAND-BINDING PROTEIN 1 PRECURSOR.
GN MLB1.
OS Streptococcus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
```

```
OC Streptococcus.
NCBI_TaxID=1306;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE C;
RX MEDLINE=96202013; PubMed=8615832;
RA Talay S.R., Grammel M.P., Chhatwal G.S.;
RT "Structure of a group C streptococcal protein that binds to
RT fibrinogen, albumin and immunoglobulin G via overlapping modules.";
RL Biochem. J. 315:577-582(1996).
DR EMBL; X84989; CAA59349.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003345; M_repeat.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02370; M; 6.
DR PRINTS; PR00015; GP0SANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Signal; Transmembrane.
FT SIGNAL 1
FT SIGNAL 36
FT SEQUENCE 439 AA; 46527 MW; D0AE6C2A85D46B89 CRC64;
SQ SEQUENCE 439 AA; 46527 MW; D0AE6C2A85D46B89 CRC64;

Query Match 19.5%; Score 151.5; DB 2; Length 439;
Best Local Similarity 27.2%; Pred. No. 0.018;
Matches 65; Conservative 25; Mismatches 46; Indels 103; Gaps 10;

Qy 19 NLGNA-----PEKALRNEERAIDELKK-----DALQSEAAVVKAD----- 57
Db 181 NLGNAKDQLDKLTTEKESLSKDEALDERKNVLEASTRTNRDLEAARDAKKAATEAELA 240
Qy 58 SSDA-----LEALADQ-----DALQSEAAVVKAD----- 83
Db 241 ETNAKVDKLEEKQVLEASRKRTRNRDLEAARDAKKAATEAELAKNELAQLEASRTNR 300
Qy 84 --NAASDALEA-----LADQTDALQSEAAVVKADNAASDALEA----- 121
Db 301 DLEAARDAKKAADAEALAKLAEALKEALKEALKEALKEALKEALKEALKEALKEAL 360
Qy 122 -----ATPIALDVKKTKDKTPV-----VKKEERONVNTLPTTGESNP 159
Db 361 GKPSMPTWGLTPATPIAKRRKT-DVKPAKANMVPYTDVKKDKR-----LPSTGETVNP 414

RESULT 7
Q54180
ID Q54180 PRELIMINARY; PRT; 103 AA.
AC Q54180;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE PROTEIN G'.
GN PROTEIN G' GENE.
OS Streptococcus sp. 'group G'.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1320;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=G148;
RX MEDLINE=90226312; PubMed=2183792;
RA Goward C.R., Murphy J.P., Atkinson T., Barstow D.A.;
RT "Expression and purification of a truncated recombinant streptococcal
RT protein G'.";
RL Biochem. J. 267:171-177(1990).
DR EMBL; X53324; CAA37409.1; -.
DR SEQUENCE 103 AA; 10828 MW; E7AE8BD1C073423C CRC64;
SQ SEQUENCE 103 AA; 10828 MW; E7AE8BD1C073423C CRC64;

Query Match 18.3%; Score 142; DB 2; Length 103;
Best Local Similarity 57.9%; Pred. No. 0.014;
Matches 33; Conservative 3; Mismatches 21; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPKALNEERAIDELKKQAEIDKEATTAEEA 57
Db 34 VDSPIEQPRIIPNGGTLNLLGNAPKALNEERAIDELKKQAEIDKEATTAEEA 90
```

```

RESULT 8
Q53900
ID Q53900 PRELIMINARY; PRT; 323 AA.
AC Q53900;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALBUMIN-BINDING PROTEIN (FRAGMENT).
OS Streptococcus canis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DG12;
RX MEDLINE=92363555; PubMed=1500168;
RA Sjobring U.;
RT "Isolation and molecular characterization of a novel albumin-binding
RL protein from group G streptococci.";
DR EMBL; M95520; AAA26847.1;
DR HSSP; Q51911; IGAB.
DR InterPro; IPR002988; GA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF01468; GA; 2.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1 323
FT NON_TER 323 323
SQ SEQUENCE 323 AA; 34990 MW; E6903CFDCD5D373E CRC64;

Query Match 17.8%; Score 138.5; DB 2; Length 323;
Best Local Similarity 33.9%; Pred. No. 0.088;
Matches 39; Conservative 25; Mismatches 36; Indels 15; Gaps 4;

QY 42 QKAEKKEATTAIE---AASSD-----ALEALADQTDALQSEEAAYVVKADNAASDALEAL 93
DB 34 ESSIKDKAQLDSYKEIKASVDRDEITELSAQADQIVSAQADNEAITKAEEDSKAWEEA 93
QY 94 ADQTDALQSEEAAYVVKADNAASDAWEKAATPIALDVKKTKDTKPVVKKKEQNVN 148
DB 94 ADQANTAKAEDELAKAEKSSDAWEKAA---ALDQAK-----QAALKEFDRYGVS 141

RESULT 9
Q91255
ID Q91255 PRELIMINARY; PRT; 1110 AA.
AC Q91255;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NF-180.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CNS;
RX MEDLINE=95287814; PubMed=7770000;
RA Jacobs A.J., Kamholz J., Selzer M.E.;
RT "The single lamprey neurofilament subunit (NF-180) lacks
RT multiphosphorylation repeats and is expressed selectively in
RT projection neurons.";
RL Brain Res. Mol. Brain Res. 29:43-52(1995).
DR EMBL; U19361; AAA80106.1;
DR InterPro; IPR001664; IF.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR PROSITE; PS00226; IF; UNKNOWN_1.
SQ SEQUENCE 1110 AA; 123818 MW; 6558DA73DAF6974C CRC64;

Query Match 15.1%; Score 117; DB 5; Length 7659;
Best Local Similarity 24.9%; Pred. No. 76;
Matches 49; Conservative 28; Mismatches 66; Indels 54; Gaps 9;

QY 9 RIIPNGGTLNLLGNA-----PEKLALR-----NEERADIELKQA- 44
DB 6812 RIAPHLATLVEAYNDVPASVEPSAVALRDRAAKFVSDLEKNIQKTGDEKRADELKNDVG 6871
QY 45 -----TED-----KEATTATAEASSDA-----LEALADQTDALQSEEAAYVVKADNA 85
DB 6872 NAVKNVEDVYVYQNPQPLDVAKDDANKLKATVEQLTKLAESSDKIDPQVAKDKDKST 6931
QY 86 AS-----DALEALADQTDALQSEEAAYVQSDNAASDAWEKAATPIALDVKKTKDTKPVVKK 141
DB 6932 KAKELQALEKAIPQEDAIRRQAEI-----NDRLNLEKELTKV--DEPKPEDALPIV-D 6984
QY 142 EERQNVNTLPTTGEEN 158
DB 142 EERQNVNTLPTTGEEN 158
```

Db 6985 QLAANTNTLKTATDSNN 7001

# RESULT 11

Q95CK5 PRELIMINARY; PRT; 166 AA.  
 AC Q95CK5;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE HYPOPHYSICAL 17.4 KDA PROTEIN.  
 GN T9C5.130.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F., Salancoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL132964; CAB62458.1; .  
 KW Hypothetical protein.  
 SQ SEQUENCE 166 AA; 17373 MW; 3C612A36544F57FE CRC64;

Query Match 14.9%; Score 116; DB 10; Length 166;  
 Best Local Similarity 26.3%; Pred. No. 1-2;  
 Matches 41; Conservative 25; Mismatches 62; Indels 28; Gaps 5;

Qy 12 PNGTTLNLGNAP-----EKLALNREERAIDELKQAIEDKEATTALAEAASSDALEALA 66  
 Db 8 PKESDMQNEGSVPKPVSENVVAKENNTGESGKQNTVAETTTSVEAKETPPVEPTK 67  
 Qy 67 DQDALQSE-----EAAV---KADNAASDALEALADQTDALQSEAEVVS 110  
 Db 68 ETPPAVQEVAVNESSADAGEAAVPAKVENAENAEKVEAVAVAAPEKVEV--- 124  
 Qy 111 DNAASDAWEKA-ATPIALDVKKTKDTPVVKKEERQ 145  
 Db 125 ---AVEAEKRAEPVKAEPVKAEPVKEESQ 157

# RESULT 12

Q53974 PRELIMINARY; PRT; 413 AA.  
 AC Q53974;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE MAG PROTEIN PRECURSOR.  
 GN MAG.  
 OS Streptococcus dysgalactiae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94259307; PubMed=7515368;  
 RA Jonsson H., Frykberg L., Rantamaki L., Guss B.;  
 RT "MAG, a novel plasma protein receptor from Streptococcus  
 RT dysgalactiae."  
 RL Gene 143:85-89(1994).  
 DR EMBL; L27798; AAA26921.1; .  
 DR HSP; P06654; IPCX  
 DR InterPro; IPR002988; GA.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR000724; Igg\_bind\_B.  
 DR Pfam; PF01468; GA; 1.

DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam; PF01378; Igg\_binding\_B; 1.  
 DR PRINTS; PR00015; GPANCHOR.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 KW Signal; Transmembrane.  
 FT SIGNAL 1 34 POTENTIAL.  
 FT CHAIN 35 413 POTENTIAL.  
 SQ SEQUENCE 413 AA; 44004 MW; F04DC71044F9E50F CRC64;

Query Match 14.7%; Score 114.5; DB 2; Length 413;  
 Best Local Similarity 26.2%; Pred. No. 4-2;  
 Matches 49; Conservative 15; Mismatches 56; Indels 67; Gaps 6;

Qy 36 AIDELKKQATD-----KEATTAIE---AASSDALEALADQTDALQSEAAVVK 81  
 Db 206 AIELKKYGGDYIKLINCKTAEGVTALKDEILASKPAVIDAPELTPALTKLVI-- 263  
 Qy 82 ADNAASDALEALADQTDALQSEEA-EVVSQSDNAASDAWE----- 119  
 Db 264 --NGKTLKGTTTKAVDAETAERAKAFKOVANENGVDGVVYDDATKTFVTETMTEVPGDA 321  
 Qy 120 -----KAATPIALDVKKTKDT-----KPVVKKEERQNVNLTPT 152  
 Db 322 PTEPKKPEASTPLVPLTPATPIAKDDAKDDTKKDDAKKPEAKKEAKKAATLPT 381  
 Qy 153 TGEESNP 159  
 Db 382 TGEESNP 388

# RESULT 13

Q935B1 PRELIMINARY; PRT; 850 AA.  
 AC Q935B1;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE PUTATIVE PHAGE TAIL PROTEIN.  
 GN HCM2.0051C.  
 OS Salmonella enterica subsp. enterica serovar Typhi.  
 OG Plasmid PHC42.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=90370;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 RA Krog A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar typhi CT18."  
 RL Nature 413:848-852(2001).  
 DR EMBL; AL513384; CAD09918.1; .  
 KW Plasmid.  
 SQ SEQUENCE 850 AA; 90917 MW; 21271493A07A6563 CRC64;

Query Match 14.7%; Score 114.5; DB 2; Length 850;  
 Best Local Similarity 31.0%; Pred. No. 9-4;  
 Matches 40; Conservative 20; Mismatches 44; Indels 25; Gaps 5;

Qy 37 IDELKKQAIEDKEATTATEAASSDAL-----EALADQTDALQSEAAVVKADNA 85  
 Db 146 VDTIESSSKALAAARAAEAAEAAALNKSQAAGSEANAKASENNAASQAARATSENA 205  
 Qy 86 ASDALEALADQTDALQSEAEVVSQSDNAASDAWEKAATPIALDVKK---TKDTPVVKKE 142





```
Db 275 ANAEKELQSOAAALEKQLEATKRELADLOAKLAATNQEKLEAEAKALKQOL-AKQVE 333
QY 104 EAEVQSDNAASDAWEKAATPIALDVYKTKDTRPVYKKEERQNVN-----TLPTTG 154
Db 334 ELAKLKADKASG-----AQKP---DTPKDNKEVTPRSQTRTNTNKAPMPOTKRQLPSTG 385
QY 155 EE-SNP 159
Db 386 EETNP 391

RESULT 17
Q95518 PRELIMINARY; PRT; 619 AA.
AC Q95518; 2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HU01392P.
GN FAU.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Y, CN BW SP.
RA Scapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celnikier S.
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY060997; AAL28945.1; -.
SQ SEQUENCE 619 AA; 68641 MW; 7DDCB26AD1AB9CEE CRC64;

Query Match 14.3%; Score 111; DB 5; Length 619;
Best Local Similarity 31.5%; Pred. No. 11;
Matches 41; Conservative 20; Mismatches 43; Indels 26; Gaps 6;

QY 4 PIEQRIIPNGTITNLGNAPEKALRNE-----EPAIDELKKQAIEDKEATTAIEAA 57
Db 297 PLEPLDV-----BAKEQRLRQLRLTVNEEALDEVD---LEKRAQKADEAK 342
QY 58 SSDALEALADOTDAL--QSEAAVVKADNAASDALEALADQ---TDALQSEAEVWSDN 112
Db 343 RREE-RALKEERDLRTAEAEKQAAKAAKAAEAAKIAEAEALLAEAAKAAEAEAKLK 401
QY 113 AASDAWEKAA 122
Db 402 AAEDAQAQKAA 411

RESULT 18
O85472 PRELIMINARY; PRT; 2055 AA.
AC O85472; 1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE EXTRACELLULAR MATRIX BINDING PROTEIN (FRAGMENT).
GN EMB.
OS Abiotrophia defectiva.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Abiotrophia.
OX NCBI_TaxID=46125;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NVS-47;
RX MEDLINE=99081722; PubMed=9864195;
RA Manganelli R., van de Rijn I.;
RT *Characterization of emb, a gene encoding the major adhesin of
```

```
RT Streptococcus defectivus.*;
RL Infect. Immun. 67:50-56(1999).
DR ENBL; AF067776; AAD03320.1; -.
FT NON_TER 2055 2055
SQ SEQUENCE 2055 AA; 215642 MW; 9699C11DD93E2FD CRC64;

Query Match 14.3%; Score 111; DB 2; Length 2055;
Best Local Similarity 25.6%; Pred. No. 43;
Matches 43; Conservative 28; Mismatches 61; Indels 36; Gaps 6;

QY 18 TNLIGNAPEKALRNEERAIDE----LKQAIED-----KEATTA 53
Db 1241 TQAINNPQTPAAKNAKAAVEQAADAKKQAIENDPNLTROEKDAATAKVQDQETNKARQA 1300
QY 54 IEAASSDALEALADOTDALQSEEA-----AVVKADNAASDALEALADOTDALQSEAEV 107
Db 1301 IDAATTNA-DVTAKONEGTQAINAVPQTPKAKTDAKNAVTOAAE---DKSAIENDPNLT 1356
QY 108 VQSDNAASDAWEKAATPI--ALDVKTKDTPVVKKEERQNVNLTPTT 153
Db 1357 REEKDAKAKVDAEATKAKNAIDAATSNDDETAKONEGTQAINAVPQT 1404

RESULT 19
Q01891 PRELIMINARY; PRT; 890 AA.
AC Q01891;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SURFACE EXCLUSION PROTEIN SEAL PRECURSOR.
GN SEAL.
OS Enterococcus faecalis (Streptococcus faecalis).
OG Plasmid pAD1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DS16;
RX MEDLINE=92293110; PubMed=1603060;
RA Weidlich G., Wirth R., Galli D.;
RT *Sex pheromone plasmid pAD1-encoded surface exclusion protein of
RT Enterococcus faecalis.*;
RL Mol. Gen. Genet. 233:161-168(1992).
DR EMBL; X62656; CAA44526.1; -.
DR PIR; S22452; S22452.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Plasmid; Signal; Membrane.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 890 SURFACE EXCLUSION PROTEIN SEAL.
SQ SEQUENCE 890 AA; 98227 MW; DCEDC2F131F634FB CRC64;

Query Match 14.2%; Score 110; DB 2; Length 890;
Best Local Similarity 29.1%; Pred. No. 19;
Matches 44; Conservative 24; Mismatches 71; Indels 12; Gaps 6;

QY 7 QPRIIPNGTITNLGNAPEKALRNEERAIDEKQAIEDKEATTAIEAASSDALEALA 66
Db 32 QPKTPENSST-----EQPTVKATQTTTEQAITE-KQQQVTEKQAIYDQKQVADTAKKEK 84
QY 67 DOTDALQSEEAAYVKADNAASD-ALEALADQTDALQSEAEVWSDNAASDAWEKAATPI 125
Db 85 DAIDQSVKQDQAVVDQNKDALDQSQAVTDQ-QAVVDEAKKV--DEATPSAIEAKEQV 141
QY 126 ALDVKTKDTPVVKKEERQNVNLTPTTGE 156
Db 142 ATDQAVDDQDKV--DQATDQVNOQAVVEE 171

RESULT 20
```

```
Q9LK53
ID Q9LK53 PRELIMINARY; PRT; 316 AA.
AC Q9LK53;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 3, P1 CLONE:WGI9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL: AP000375; BAB01411.1; -.
SQ SEQUENCE 316 AA; 35639 MW; 743F0166E7E93230 CRC64;

Query Match 14.0%; Score 109; DB 10; Length 316;
Best Local Similarity 27.3%; Pred. No. 7;
Matches 48; Conservative 21; Mismatches 51; Indels 56; Gaps 7;

QY 27 KLALNEERAIDELKK-----QAIED-----KEATTAIEAASDALEALADQTDALQ 73
DB 132 KLELONVERKAHSELKEIQERDHOAIEDLKKETDKATQLSLEELXIAVFEAQEAKD 191
QY 74 SEE-----AAVVKADNA-----ASDALEALADQTDALQSEAEVQSD 111
DB 192 AEEHARLNVAVLESDFRS LAVVRESAAEELTFETALRACRDETLKTMESEREIEDIK 251
QY 112 NAASDAWEKA-----ATPIALDVKK-----TKDTPVVKKEERQN 146
DB 252 AATQDALKKAEMAEQATIIVVDVLRKRRRKAASRIILAESKMKCAKSTKEVLKSRPRS 307

RESULT 21
Q54860 PRELIMINARY; PRT; 425 AA.
AC Q54860;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE IMMUNOGLOBULIN-FC-BINDING PROTEIN.
GN FCRA2.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M2;
RX MEDLINE=94261050; PubMed=8202029;
RA Podbielski A., Weber-Heynemann J., Cleary P.P.;
RT "Immunoglobulin-binding FerA, Enn proteins and M proteins of group A
RT Streptococci, evolved independently from a common ancestral protein.";
RL Med. Microbiol. Immunol. 183:33-42(1994).
DR EMBL: X73159; CAA51682.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PRINTS: PR00015; GP05ANCHOR.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Transmembrane.

Q9LK53
SQ SEQUENCE 425 AA; 46194 MW; CBE382280190D0CD CRC64;
Query Match 14.0%; Score 109; DB 2; Length 425;
Best Local Similarity 26.6%; Pred. No. 9.8;
Matches 53; Conservative 27; Mismatches 25; Indels 60; Gaps 10;

QY 6 EQPRIIPNGGTITNLLGNA-----PEKLAIRNEERAIDEL---KKOAIED 47
DB 218 ELAKLQSEAAATLENLLGSAKRELTELQAKLDTATAEKAKLESQVTTLENLLGSAKRELTD 277
QY 48 KEATTAIEAASDALEB-----ALADQTDALQSEEA---AVVKADNAASDALE---ALADQ 96
DB 278 LQAKLDAANAEXEKQLQSQAAALEKQLKLEATKELADLQAKLAATNQGKEKLEAEAKALKDQ 337
QY 97 TIALQSEAEVQSDNAASDAWEKAATPIALDVKKTKDKPVVKK-----EERQNVN-- 148
DB 338 L-AKQAEELAKUKAKKAGSG-----QKPDTPKGNKEVTPRPSQTRTNTNKA 382
QY 149 -----TLPTTGE--SNP 159
DB 383 PMAQTKRQLPSTGEETNP 401

RESULT 22
Q54862 PRELIMINARY; PRT; 159 AA.
ID Q54862;
AC Q54862;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE DECLINED PROTEIN DURING SEED DEVELOPMENT.
OS Fagopyrum esculentum (Common buckwheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Polygonaceae; Fagopyrum.
OX NCBI_TaxID=3617;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. KITAYUKI; TISSUE=IMMATURE SEED;
RA Fujino K., Funatsuki H., Inada M., Shimono Y., Kikuta Y.;
RT "Expression and sequence analysis of cDNAs corresponding to genes
RT differentially expressed during seed development of buckwheat
RT (Fagopyrum esculentum Moench).";
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: D87983; BAA21761.1; -.
SQ SEQUENCE 159 AA; 16656 MW; 56862A34D99E607E CRC64;

Query Match 13.9%; Score 108; DB 10; Length 159;
Best Local Similarity 30.1%; Pred. No. 3.8;
Matches 41; Conservative 26; Mismatches 47; Indels 22; Gaps 6;

QY 12 PNCGLTITNLLGNAPEKLAIRNEERAIDELKKOAIEDKEATTAIEAASDALEALADQTD 71
DB 8 PN---TTTLOENEKSEVAQVEEVAQVEETPAVE---EAPAAVEEAPAAVEEVEEAPA 62
QY 72 LQSEAAVVKADNAASDALEALADQTDALQSEAEVQSDNAASDAWEKAATPIALDVKK 131
DB 63 ATEEAPAAVE-EAPAAVEEAVA-----ETKEVVEEVEEKAA---EAPTPEEKEKK 110
QY 132 TKDTPVVKKEERQNV 147
DB 111 VEEV-----EEKQV 121

RESULT 23
Q54862 PRELIMINARY; PRT; 405 AA.
ID Q54862;
AC Q54862;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FC-RECEPTOR PROTEIN PRECURSOR (FRAGMENT).
GN FCRA76.
```

```

OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M TYPE 76;
RX MEDLINE=89282846; PubMed=2660147;
RA Heath D.G., Cleary P.P.;
RT *Fc-receptor and M-protein genes of group A Streptococci are products
RT of gene duplication.;
RL Proc. Natl. Acad. Sci. U.S.A. 86:4741-4745(1989).
DR EMBL; M22532; AAB95296.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Signal; Receptor.
FT SIGNAL 1 41 POTENTIAL.
FT CHAIN 42 >405 FC-RECEPTOR PROTEIN.
FT NON_TER 405
SQ SEQUENCE 405 AA: 44299 MW: 9B73D62D515B41F3 CRC64;

Query Match 13.9%; Score 108; DB 2; Length 405;
Best Local Similarity 27.5%; Pred. No. 11;
Matches 53; Conservative 25; Mismatches 67; Indels 48; Gaps 10;

QY 6 EQPRIIPNGGTLTNLIGNA-----PEKLALNEERAIDEL---KKQAIED 47
DB 218 ETAKLOSEAATLENLIGSAKRELTDLOAKLDATAEAKAKLESQVTTLENLIGSAKRELTD 277
QY 48 KEATTAIEAASSDALE----ALADQTDALQSEEA---AVVKADNAASDALE----ALADQ 96
DB 278 LQAKLDAANAEEKLQSQAAALEKQLEATKKELADLOAKLAATNOEKEKEAEAKALKQ 337
QY 97 TDALQSEEAEEVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKKEERQNVN----- 148
DB 338 L-AKQAEELAKLKADKASG-----AQKP---DTKPCNKEVPTPSPQTRTNTNKPAPQTK 388
QY 149 -TLPTTGEE-SNP 159
DB 389 RQLPSTGEETTNP 401

RESULT 24
P95813
ID P95813 PRELIMINARY; PRT; 415 AA.
AC P95813;
DT 01-MAY-1997 (TRENBLrel. 03, Created)
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE FC-GAMMA RECEPTOR PRECURSOR.
GN FCRA9.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=71-683;
RX MEDLINE=96396849; PubMed=8803952;
RA Krebs B., Kaufhold A., Boyle M.D.P., Podbielski A.;
RT *Different alleles of the forA/mrp gene of Streptococcus pyogenes
RT encode M-related proteins exhibiting an identical immunoglobulin-
RT binding pattern.;
RL Med. Microbiol. Immunol. 185:39-47(1996).
DR EMBL; X78483; CAA55229.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GP0SANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DE FC-GAMMA RECEPTOR PRECURSOR.
GN FCRA9.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=71-683;
RX MEDLINE=96396849; PubMed=8803952;
RA Krebs B., Kaufhold A., Boyle M.D.P., Podbielski A.;
RT *Different alleles of the forA/mrp gene of Streptococcus pyogenes
RT encode M-related proteins exhibiting an identical immunoglobulin-
RT binding pattern.;
RL Med. Microbiol. Immunol. 185:39-47(1996).
DR EMBL; X78482; CAA55229.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GP0SANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Receptor; Signal; Transmembrane.
FT SIGNAL 1 41 POTENTIAL.
FT CHAIN 42 >405 FC-RECEPTOR PROTEIN.
FT NON_TER 405
SQ SEQUENCE 415 AA: 45256 MW: B4C67D8E230E56DF CRC64;

```

```

Query Match 13.9%; Score 108; DB 2; Length 415;
Best Local Similarity 27.5%; Pred. No. 11;
Matches 53; Conservative 25; Mismatches 67; Indels 48; Gaps 10;

QY 6 EQPRIIPNGGTLTNLIGNA-----PEKLALNEERAIDEL---KKQAIED 47
DB 208 ETAKLOSEAATLENLIGSAKHELTDLOAKLDATAEAKAKLESQVTTLENLIGSAKRELTD 267
QY 48 KEATTAIEAASSDALE----ALADQTDALQSEEA---AVVKADNAASDALE----ALADQ 96
DB 268 LQAKLDAANAEEKLQSQAAALEKQLEATKKELADLOAKLAATNOEKEKEAEAKALKQ 327
QY 97 TDALQSEEAEEVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKKEERQNVN----- 148
DB 328 L-AKQAEELAKLKADKASG-----AQKP---DTKPCNKEVPTPSPQTRTNTNKPAPQTK 378
QY 149 -TLPTTGEE-SNP 159
DB 379 RQLPSTGEETTNP 391

RESULT 25
P95808
ID P95808 PRELIMINARY; PRT; 425 AA.
AC P95808;
DT 01-MAY-1997 (TRENBLrel. 03, Created)
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE FC-GAMMA RECEPTOR PRECURSOR.
GN FCRA13.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=71-686;
RX MEDLINE=96396849; PubMed=8803952;
RA Krebs B., Kaufhold A., Boyle M.D.P., Podbielski A.;
RT *Different alleles of the forA/mrp gene of Streptococcus pyogenes
RT encode M-related proteins exhibiting an identical immunoglobulin-
RT binding pattern.;
RL Med. Microbiol. Immunol. 185:39-47(1996).
DR EMBL; X78483; CAA55232.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GP0SANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Receptor; Signal; Transmembrane.
FT SIGNAL 1 41 POTENTIAL.
FT CHAIN 42 >405 FC-RECEPTOR PROTEIN.
FT NON_TER 405
SQ SEQUENCE 425 AA: 46558 MW: 1DD73FB2CF76DAD8 CRC64;

Query Match 13.9%; Score 108; DB 2; Length 425;
Best Local Similarity 27.5%; Pred. No. 11;
Matches 53; Conservative 25; Mismatches 67; Indels 48; Gaps 10;

QY 6 EQPRIIPNGGTLTNLIGNA-----PEKLALNEERAIDEL---KKQAIED 47
DB 218 ETAKLOSEAATLENLIGSAKHELTDLOAKLDATAEAKAKLESQVTTLENLIGSAKRELTD 277
QY 48 KEATTAIEAASSDALE----ALADQTDALQSEEA---AVVKADNAASDALE----ALADQ 96
DB 278 LQAKLDAANAEEKLQSQAAALEKQLEATKKELADLOAKLAATNOEKEKEAEAKALKQ 337
QY 97 TDALQSEEAEEVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKKEERQNVN----- 148
DB 338 L-AKQAEELAKLKADKASG-----AQKP---DTKPCNKEVPTPSPQTRTNTNKPAPQTK 388
QY 149 -TLPTTGEE-SNP 159
DB 389 RQLPSTGEETTNP 401

```



<hr/>					
Query Match	13.9%	Score 108;	DB 3;	Length 10367;	
Best Local Similarity	21.8%;	Pred. No. 33;			
Matches	36;	Conservative	42;	Mismatches	61; Indels 26; Gaps
<hr/>					
QY	11	IPNGGTLTNLGNAPKALRNE-----ERAIDELKKQAIEDKEATTAAEAS----	SD 60		
DB	874	VPN-----INSPEKMVDSEADEKASKAEVQIKKAELNDOEEI KEDSPHSD	926		
<hr/>					
QY	61	ALEALADOTDALQSEEAAYVKADNAASDALEALADOTDALQSEEAFFVWS-DNAASDAWE	119		
DB	927	SVAELREAVTASEGDVVHLEADNRKIRKELLDELIAEAK-PDQEKDIVOSENATTEALD	985		
<hr/>					
QY	120	KAAF-----PTALDVKTCTKTPVVKKEERQNNTLTPTTGES	157		
DB	986	RSVTDEDDLVPPTQVSQEIMEIDEPAEAKAANNENRTTRADEKEA	1030		
<hr/>					
RESULT 30					
ID	P95810	PRELIMINARY;	PRT;	384 AA.	
AC	P95810;				
DT	01-MAY-1997 (TrEMBLrel. 03, Created)				
DT	01-MAY-1997 (TrEMBLrel. 03, Last sequence update)				
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)				
DT	FCRAB PROTEIN PRECURSOR.				
GN	FCRAB.				
OS	Streptococcus pyogenes.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				
OX	Streptococcus.				
NCBI_TaxID=1314;					
RN	[1]				
RC	SEQUENCE FROM N.A.				
RP	STRAIN=M8-4025;				
RA	MEDLINE=96396849; PubMed=8803952;				
RX	Krebs B., Kaufhold A., Boyle M.D.P., Podbielski A.;				
RT	"different alleles of the fcrA/mrp gene of Streptococcus pyogenes				
RT	encode M-related proteins exhibiting an identical immunoglobulin-				
RL	binding pattern.";				
RL	Med. Microbiol. Immunol. 185:39-47(1996).				
DR	EMBL; X78484; CAA35235.1; -				
DR	InterPro: IPR001899; Gram_pos_anchor.				
DR	Pfam; PF00746; Gram_pos_anchor; 1.				
DR	PRINTS; PR00015; GP0SANCHOR.				
DR	DROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.				
KW	Signal; Transmembrane.				
FT	SIGNAL	1	41	POTENTIAL.	
SEQ	SEQUENCE	384 AA;	41708 MW;	8DF120B6AA30BAD5 CRC64;	
<hr/>					
Query Match	13.8%;	Score 107;	DB 2;	Length 384;	
Best Local Similarity	28.4%;	Pred. No. 12;			
Matches	52;	Conservative	23;	Mismatches	60; Indels 48; Gaps
<hr/>					
QY	16	TFLNLGNA-----PEKLALRNPERAIDEL---KKQAIEDKEATTAIEAA	57		
DB	187	TLENLLGSARKRELTDLOAKLDATAEAKLESQVTTLENLLGSARKRELTDLOAKLDAANA	246		
<hr/>					
QY	58	SSDALE----ALADQTDALQSEFA---AVVKADNAASDALE----ALADQTDALQSEFAE	106		
DB	247	EKEKLQSAAALEKQLATEKRELADLOAKLAATNQEKLEAFALKKEOL-AKQAEELA	305		
<hr/>					
QY	107	VVGSDNAASDAWEKAATPTALDVKTKDKTPVVKKEERQNVN-----TLPTTGEE-	156		
DB	306	KLADRASG-----AQKP---DTKPGNKVEPTRPSQTRTNKAPMAQTKRQLPSTGEET	357		
<hr/>					
QY	157	SNP 159			
DB	358	TNP 360			
<hr/>					
RESULT 31					
ID	Q53474	PRELIMINARY;	PRT;	388 AA.	
ID	Q53474				
AC	Q53474;				

DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE MRP4.  
 GN PCRA15.  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95147851; PubMed=7845360;  
 RA Katerov V., Schalen C., Totolian A.A.;  
 RT "Sequencing of genes within the vir regulon of Streptococcus pyogenes  
 RT type M15-an opacity factor-positive serotype with low opacity factor  
 RT expression.";  
 RL Mol. Gen. Genet. 245:78-85(1994).  
 DR EMBL; S75411; AAB33261.1; -.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR PRINTS; PR00015; GPOSANCHOR.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 KW Transmembrane.  
 SQ SEQUENCE 388 AA; 41957 MW; 2C7320D575A5C899 CRC64;

Query Match 13.8%; Score 107; DB 2; Length 388;  
 Best Local Similarity 30.4%; Pred. No. 12;  
 Matches 49; Conservative 18; Mismatches 55; Indels 39; Gaps 8;

QY 16 TLTNLLGNAP-----KLALNEERAIDELKKQAIEDK-----EATTA-----TEAASSDA 68  
 DB TLTLNLLGNAP-----KLALNEERAIDELKKQAIEDK-----EATTA-----TEAASSDA 68  
 DB 226 TLNLLGSAKRELTDQAKLDAANAEEA--KLQSA-----AALEKQLEATKKE 272  
 QY 69 TDAQSEAAVKNADNAASDALEALADQTDALQSEAEVVSQSDAASDAWEKAATPTALD 128  
 DB TLTLNLLGSAKRELTDQAKLDAANAEEA--KLQSA-----AALEKQLEATKKE 272  
 DB 273 LADLQAKLAATNQEKELEAKALEQL-AKQAELEAKLADKASG-----AKRP---D 323  
 QY 129 VKYTKDKTPVVKKEQRNVN-----TLPTTGEE-SNP 159  
 DB 324 TKPGNEVPTPSQTRNTNKAPMAQTKQLPSTGETNP 364

RESULT 32  
 Q9EYEB ID Q9EYEB PRELIMINARY; PRT; 437 AA.  
 AC Q9EYEB;  
 DT 01-WAR-2001 (TREMBLrel. 16, Created)  
 DT 01-WAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE TAIL FIBER PROTEIN.  
 GN H0004.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=O157:H7;  
 RC MEDLINE=20564182; PubMed=11111050;  
 RA Yokoyama K., Makino K., Kubota Y., Watanabe M., Kimura S.,  
 RA Yutsudo C.H., Kurokawa K., Ishii K., Hattori M., Abe H., Iida T.,  
 RA Yamamoto K., Hayashi T., Yasunaga T., Honda T., Sasakawa C.,  
 RA Shinagawa H.;  
 RT "Complete nucleotide sequence of the defective Sakai-VT1 prophage  
 RT carrying the shiga toxin 1 genes of the enterohemorrhagic Escherichia  
 RT coli O157:H7 derived from the Sakai outbreak.";  
 RL Gene 258:127-139(2000).  
 DR EMBL; AF000400; BAB19560.1; -.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01391; Collagen; 2.  
 SQ SEQUENCE 437 AA; 44183 MW; 408C870ACD31E14D CRC64;

Query Match 13.7%; Score 106.5; DB 2; Length 437;

Best Local Similarity 28.5%; Pred. No. 15;  
 Matches 47; Conservative 24; Mismatches 67; Indels 27; Gaps 7;

QY 15 GTLTNLLGNAP-----KLALNEERAIDELKKQAIEDK-----EATTA-----TEAASSDA 61  
 DB GTLTNLLGNAP-----KLALNEERAIDELKKQAIEDK-----EATTA-----TEAASSDA 61  
 DB 84 GTLNDFLGANSESDVRPEALRFEELVVEAARHAEAKKNAGEAETSARNAGISASQAE 143  
 QY 62 LEALADQT--DALQS-----EEAAVVK-ADNAASDALEALADQTDALQSEAEVVSQSDA 113  
 DB LEALADQT--DALQS-----EEAAVVK-ADNAASDALEALADQTDALQSEAEVVSQSDA 113  
 DB 144 SAANADTSAGDALESARQAASAAAKQSDASSSSASAAQAASQSSQSAAEISRK 203  
 QY 114 ASDAWEKAATPTIALDVKTKDKTPVVKKEER-----QNVNTLPT 152  
 DB ASDAWEKAATPTIALDVKTKDKTPVVKKEER-----QNVNTLPT 152  
 DB 204 AESAAGNAARDATTATEKARESAESQAESQSRIAAEAVNRPT 248

RESULT 33  
 P71213 ID P71213 PRELIMINARY; PRT; 314 AA.  
 AC P71213;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE ESPB GENE (SECRETED PROTEIN ESPB).  
 GN ESPB.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=STEC;  
 RX MEDLINE=97045129; PubMed=8890194;  
 RA Ebel F., Delbel C., Kresse A.U., Guzman C.A., Chakraborty T.;  
 RT "Temperature- and medium-dependent secretion of proteins by Shiga  
 RT toxin-producing Escherichia coli.";  
 RL Infect. Immun. 64:4472-4479(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RDEC-1;  
 RX MEDLINE=97427930; PubMed=9284118;  
 RA Abe A., Kenny B., Stein M., Finlay B.B.;  
 RT "Characterization of two virulence proteins secreted by rabbit  
 RT enteropathogenic Escherichia coli, EspA and EspB, whose maximal  
 RT expression is sensitive to host body temperature.";  
 RL Infect. Immun. 65:3547-3555(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RDEC-1;  
 RA Abe A., Kenny B., Stein M., Finlay B.B.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 8-308 FROM N.A.  
 RC STRAIN=85/150;  
 RA Abe A., Heczko U., Hegge R.G., Finlay B.B.;  
 RT "Two Enteropathogenic Escherichia coli Type III Secreted Proteins,  
 RT EspA and EspB, are Virulence Factors.";  
 RL J. Exp. Med. 0:0-0(1998).  
 RN [5]  
 RP SEQUENCE OF 8-308 FROM N.A.  
 RC STRAIN=85/150;  
 RA Lau G., Abe A., Finlay B.B.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B10;  
 RX MEDLINE=99141579; PubMed=9987106;  
 RA Nougayrede J.P., Marches O., Boury M., Mainil J., Charlier G.,  
 RA Pohl P., De Rycke J., Milon A., Oswald E.;  
 RT "The long-term cytoskeletal rearrangement induced by rabbit  
 RT enteropathogenic Escherichia coli is Esp dependent but intimin  
 RT independent.";  
 RL Mol. Microbiol. 31:19-30(1999).

```
[7]
RP SEQUENCE FROM N.A.
RC STRAIN=565/56, 84/110-1, AND 83/39;
RA Krejany E.O.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
[8]
RP SEQUENCE FROM N.A.
RC STRAIN=RDEC-1;
RX MEDLINE=98254123; PubMed-9593291;
RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;
RT "The complete sequence of the locus of enterocyte effacement (LEE)
from enteropathogenic Escherichia coli E2348/69.";
RL Mol. Microbiol. 28:1-4(1998).
[9]
RP SEQUENCE FROM N.A.
RC STRAIN=RDEC-1;
RX MEDLINE=21153569; PubMed-11254564;
RA Zhu C., Agin T.S., Elliott S.J., Johnson L.A., Thate T.E., Kaper J.B.,
RA Boedeker E.C.;
RT "Complete Nucleotide Sequence and Analysis of the Locus of Enterocyte
Effacement from Rabbit Diarrheagenic Escherichia coli RDEC-1.";
RL Infect. Immun. 69:2107-2115(2001).
DR EMBL; X99670; AAC67984.1; -
DR EMBL; U80796; AAB69980.1; -
DR EMBL; AF059713; AAC64002.1; -
DR EMBL; AF054421; AAC82360.1; -
DR EMBL; AF144010; AAD34584.1; -
DR EMBL; AF144008; AAD34582.1; -
DR EMBL; AF144009; AAD34583.1; -
DR EMBL; AF200363; AAK26729.1; -
SQ SEQUENCE 314 AA; 33220 MW; D9B781454CCA5267 CRC64;

Query Match 13.6%; Score 105.5; DB 2; Length 314;
Best Local Similarity 28.7%; Pred. No. 12;
Matches 45; Conservative 23; Mismatches 68; Indels 21; Gaps 6;

QY 14 GGTTLNLLGNAPEKALRNEERAIDEKQA--IEDEKATP---ATEAASDALEALADQ 68
DB 100 GGAISVVG-----ILGSFAINSATKGSADIAQKTASTSSKADDAASDTATKLTKA 152

QY 69 T-----DALQSEEAHVKNADNAADALEALADQTDALQSEEAHVQSDNAASDAWEKAATP 124
DB 153 TESVADAVEADASSVMOQAMTTATRAASRTSDVADDIADSAQSAQSLAENADAAQAKRAS 212

QY 125 ----IALDVKKTKDTRKPVVKKKEERQNVTLPTTGEES 157
DB 213 SRFMAAVD-KITGSTPTPIAVTSIAEGTKTLPTTVSES 248

RESULT 34
Q92IX5 PRELIMINARY; PRT; 389 AA.
AC Q92IX5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE WP22.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AL168;
RA Thern A., Mustfelt M., Lindahl G.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007557; AAD01409.1; -
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Transmembrane.
```

```
SQ SEQUENCE 389 AA; 42424 MW; 68F2EC0FCF15C5EC CRC64;

Query Match 13.6%; Score 105.5; DB 2; Length 389;
Best Local Similarity 27.4%; Pred. No. 15;
Matches 54; Conservative 26; Mismatches 68; Indels 49; Gaps 11;

QY 3 SPIEQ-PRIPNGGTLTNLLGNA-----PEKIALRNEERAIDEI---KKQ 43
DB 178 SKIEETAKLQSEAAATLENLLGSAKRELTDQAKLDTATAEKAKLESRTVTLENLLGSAKR 237
QY 44 AIEDKATTAIEAASDALEA---LADQTDALQSEEA---AVVKADNAASDALE---A 92
DB 238 ELTDQAKLDASNAEKEKLSQQAATLEKQLEATKKELADLQAKLAATNOEKELEAEAKA 297
QY 93 LADQTDALQSEEAHVQSDNAASDAWEKAATPIALDVKKTKDTRKPVVKKKEERQNVN--- 148
DB 298 LKSQL-AKQAEELAKIKADKASG-----AQKP---DTKPGNKEVTPRPSQTRINTNKAPM 348
QY 149 -----TLPTTGEES-SNP 159
DB 349 AQTAKRLPSTGEETNP 365

RESULT 35
Q54842 PRELIMINARY; PRT; 387 AA.
AC Q54842;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FCRA64/14 PROTEIN PRECURSOR.
GN FCRA64/14.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=64;
RX MEDLINE=94178942; PubMed-8132341;
RA Boyle M.D.P., Hawlitzky J., Raeder R., Podbielski A.;
RT "Analysis of genes encoding two unique type Iia immunoglobulin G-
RT binding proteins expressed by a single group A streptococcal
RT isolate.";
RL Infect. Immun. 62:1336-1347(1994).
DR EMBL; X72932; CAA51436.1; -
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Signal; Transmembrane.
FT SIGNAL 1 41 POTENTIAL.
FT CHAIN 42 387 POTENTIAL.
SQ SEQUENCE 387 AA; 42153 MW; 5989493B3D4AD5EA CRC64;

Query Match 13.5%; Score 105; DB 2; Length 387;
Best Local Similarity 27.5%; Pred. No. 16;
Matches 53; Conservative 25; Mismatches 67; Indels 48; Gaps 10;

QY 6 EOPRIIPNGGTLTNLLGNA-----PEKIALRNEERAIDEI---KKQAIED 47
DB 180 EIAKLQSEAAATLENLLGSAKRELTDQAKLDTATAEKAKLESQVTTLENLLGSAKRELTD 239
QY 48 KEATTATEAASDALE---ALADQTDALQSEEA---AVVKADNAASDALE---ALADQ 96
DB 240 LQAKLDRAAEKEKLSQQAALKEKLEATKKELADLQAKLVATNOEKELEAEAKALKEQ 299
QY 97 TDALQSEEAHVQSDNAASDAWEKAATPIALDVKKTKDTRKPVVKKKEERQNVN----- 148
DB 300 L-AKQAEELAKIKADKASG-----AQKP---DTKPGNKEVTPRPSQTRINTNKASMAQTK 350
QY 149 -TLPTTGEES-SNP 159
DB 11:1111:11
```

```
Db 351 RELPSTGEETNP 363
RESULT 36
Q3K2J9
ID Q9K2J9 PRELIMINARY; PRT; 524 AA.
AC Q9K2J9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SECRETED ANTIGEN SAGA.
GN SAGA.
OS Enterococcus faecium (Streptococcus faecium).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1352;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DO;
RA Teng F., Murray B.E., Weinstock G.M.;
RT "Characterization of an enterococcal secreted antigen.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN [2]
RP SEQUENCE FROM N.A.
RA Teng F., Murray B.E., Weinstock G.M.;
RT "Characterization of an enterococcal secreted antigen.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF167575; AAF87092.1; -.
DR EMBL; AF242195; AAF86217.1; -.
DR MEROPS; C40.0PW; -.
DR InterPro; IPR000064; NLP_C_P60.
DR Pfam; PF00877; NLP_C_P60; 1.
SQ SEQUENCE 524 AA; 55494 MW; E7BE1A611FAD9569 CRC64;

Query Match 13.5%; Score 105; DB 2; Length 524;
Best Local Similarity 23.6%; Pred. No. 22;
Matches 33; Conservative 31; Mismatches 66; Indels 10; Gaps 3;

QY 26 EKALNRERADLKKQAIEDKATTALEAASSDALEALADOTDALQSEEAAYVKAADNA 85
Db 227 EQARREQRRLAEARQQNAQKAEKEAQAEAEQAQASST---AQSATEESSAT 282
QY 86 ASDALEALADOTDALQSEEAAYVQSDNAADWEKAAT---PIALDVKTKDTKPVVKE 142
Db 283 QSSMTEESSATQSSATEESTTPESSTESTAPESSTESTAPESSTESTTPVPESS 342
QY 143 ERQNVNTPD---TTGEESNP 159
Db 343 ATEESTTPESSTTEESTTP 362

RESULT 37
Q33975
ID Q33975 PRELIMINARY; PRT; 664 AA.
AC Q33975;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE IMMUNOGLOBIN G BINDING PROTEIN MIG PRECURSOR (IGG BINDING PROTEIN MIG).
GN MIG.
OS Streptococcus dysgalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1334;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=SC1;
RA MEDLINE=94192673; PubMed=8143736;
RA Jonsson H., Mueller H.-P.;
RT "The type-III Fc receptor from Streptococcus dysgalactiae is also an alpha 2-macroglobulin receptor.";
RL Eur. J. Biochem. 220:819-826(1994).
```

```
RN [2]
RP FUNCTION.
RX MEDLINE=99081771; PubMed=9864244;
RA Vasi J., Svensson J., Frick I.-M., Mueller H.-P.;
RT "Five homologous repeats of the protein G-related protein MIG cooperate in binding to goat immunoglobulin G.";
RL Infect. Immun. 67:413-416(1999).
CC 1- FUNCTION: BINDS TO THE CONSTANT FC REGION OF IGG WITH HIGH AFFINITY. N-TERMINUS MEDIATES BINDING TO PLASMA PROTEINASE INHIBITOR ALPHA 2-MACROGLOBULIN AFTER COMPLEX FORMATION WITH PROTEASES.
CC 1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL (BY SIMILARITY).
CC 1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS IN THE REGION OF THE MEMBRANE ANCHOR. C-TERMINAL IS SIMILAR TO OTHER STREPTOCOCCAL TYPE-III FC RECEPTORS.
DR EMBL; Z29666; CAA82764.1; -.
DR HSSP; P06654; IPGX.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000724; Igg_bind_B.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF01378; Igg_binding_B; 5.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC.1; UNKNOWN.1.
KW IGG-binding protein; Repeat; Transmembrane; Cell wall; Signal.
FT SIGNAL 1 30
FT CHAIN 31 664
FT DOMAIN 31 208
FT DOMAIN 559 638
FT TRANSMEM 639 659
FT DOMAIN 660 684
FT DOMAIN 87 189
FT REPEAT 87 110
FT REPEAT 166 189
FT DOMAIN 209 558
FT REPEAT 209 278
FT REPEAT 279 348
FT REPEAT 349 418
FT REPEAT 419 488
FT REPEAT 489 558
FT DOMAIN 592 621
FT DOMAIN 630 635
SQ SEQUENCE 664 AA; 72023 MW; 4A4LCBEF7977862A CRC64;

Query Match 13.5%; Score 105; DB 2; Length 664;
Best Local Similarity 23.9%; Pred. No. 29;
Matches 54; Conservative 21; Mismatches 75; Indels 76; Gaps 9;

QY 1 VDSPIEQP-----RIIPNGGTLTNLIGNAPEKLALNRERADLKKQAIED----- 47
Db 423 IDAPELTALTYTKLVKNTFS---GETTKAV--DAETAKEAFKOYANENVYCEVSY 477
QY 48 KEATTAIEAASDALEALADOTDALQSEEAAYVKADNAASDALEALADOTDALQSEEA-E 106
Db 478 DDATKTFVTVEKPAVIDAPELTALTYTKLVF----NGTKLGETTTKAVDAETAKEAFK 533
QY 107 VVQSDNNAASDAW-----EK-----AATPIALDGV 129
Db 534 QVANENGVDGVWYDDATKTFVTVMVTPGDATPEPEKPEASIPLVPLTPTATPIAKDD 593
QY 130 KTKDTKPV-----VKKERQNVNTPPTTGEESNP 159
Db 594 AKKDDTKKVDTKKEDAKKPEAKKEAKKAATLPTTGEESNP 639

RESULT 38
Q47802
ID Q47802 PRELIMINARY; PRT; 843 AA.
```



```
AC Q47802; Q52199;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (T-EMBLrel. 15, Last annotation update)
DE SURFACE EXCLUSION PROTEIN PRECURSOR (SEAL).
GN SEPI.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Plasmid pp01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97074879; PubMed=8917306;
RA Hirt H., Wirth R., Muscholl A.;
RT "Comparative analysis of 18 sex pheromone plasmids from Enterococcus
RT faecalis: detection of a new insertion element on pp01 and
RT implications for the evolution of this plasmid family.";
RL Mol. Gen. Genet. 252:640-647(1996).
RN [2]
RP SEQUENCE OF 1-127 FROM N.A.
RX MEDLINE=96032393; PubMed=7559344;
RA Nakayama J., Yoshida K., Kobayashi H., Isogai A., Clewell D.B.,
RA Suzuki A.;
RT "Cloning and characterization of a region of Enterococcus faecalis
RT plasmid pp01 encoding pheromone inhibitor (ipd), pheromone sensitivity
RT (traC), and pheromone shutdown (traB) genes.";
RL J. Bacteriol. 177:5567-5573(1995).
RN [3]
RP SEQUENCE OF 1-127 FROM N.A.
RX MEDLINE=96032394; PubMed=7559345;
RA Fujimoto S., Tomita H., Wakamatsu E., Tanimoto K., Ike Y.;
RT "Physical mapping of the conjugative bacteriocin plasmid ppD1 of
RT Enterococcus faecalis and identification of the determinant related to
RT the pheromone response.";
RL J. Bacteriol. 177:5574-5581(1995).
DR EMBL; X96976; CAA65662.1; -.
DR EMBL; D78016; BAA11202.1; -.
KW Signal; Plasmid.
FT SIGNAL. 1 30 POTENTIAL.
FT SEQUENCE 843 AA; 93049 MW; 632C8BDC8E79EB2C CRC64;
OS Streptococcus suis.
Query Match 13.4%; Score 104.5; DB 2; Length 843;
Best Local Similarity 28.3%; Pred. No. 41;
Matches 39; Conservative 27; Mismatches 61; Indels 11; Gaps 5;

QY 7 OPRIPNGTILNLLGNAPEKALRNEERAIDELKKQAIEDKEATTATAEASDALEALA 66
Db 32 QPK-----TPENASSEQPTVKATQTTEQAITEKQOOVIE-KQAIVDKQKQVADTAKKEK 84

QY 67 DQTD-ALQSEEAHVKADNAASDALEALADQTDALQSEEAHVWQSDNAASDAWEKAATPI 125
Db 85 DTIDQSVKQAVVDQNKDALVQSQAVTQ-QAVVDEAKV--DEATPSAIEKAQEV 141

QY 126 ALDVKTKTKTPVVKKEE 143
Db 142 ATDQAVDDQKQVVEQAQ 159

RESULT 39
Q9BHC4 PRELIMINARY; PRT; 924 AA.
AC Q9BHC4;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE POSSIBLE R27-2 PROTEIN (FRAGMENT).
GN P883.45.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=FRIEDLIN;
RA Saunders D., Murphy L., Silva W., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL590734; CAC37243.1; -.
FT NON_TER 924
SQ SEQUENCE 924 AA; 99390 MW; 28F42FB20499E85A CRC64;
Query Match 13.4%; Score 104.5; DB 5; Length 924;
Best Local Similarity 25.9%; Pred. No. 46;
Matches 37; Conservative 24; Mismatches 45; Indels 37; Gaps 4;

QY 5 IEQPRIIPNGTILNLLGNAPEKALRNEERAIDELKKQAIEDKEATTATAEAA 57
Db 484 VEEMRSLISGLTLTLHLQVTDKAGDVPPEAEAPQPSQLERSLSTEEVADQGTSAVDAA 543

QY 58 SSDALEALADQTDALQSEEAHVKADNAASDALEA-----LADQTDAL 100
Db 544 SSEA-----QKEAALAEAAADEPAEATKWSVLQGEIDNLADSIADFLKQ 592

QY 101 OSEEA--EVVOSDNAASDAWEKA 121
Db 593 HSOEATLEATASOKADAPAVEEA 615

RESULT 40
Q07290 PRELIMINARY; PRT; 1822 AA.
AC Q07290;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (T-EMBLrel. 15, Last annotation update)
DE EF PROTEIN.
GN EPF*.
OS Streptococcus suis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1307;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1890;
RX MEDLINE=93328288; PubMed=8333363;
RA Smith H.E., Reek F.H., Vecht U., Gielkens A.L.J., Smits M.A.;
RT "Repeats in an extracellular protein of weakly pathogenic strains of
RT Streptococcus suis type 2 are absent in pathogenic strains.";
RL Infect. Immun. 61:3318-3326(1993).
DR EMBL; X71880; CAA50714.1; -.
SQ SEQUENCE 1822 AA; 192631 MW; 3838960C77641D7D CRC64;
Query Match 13.4%; Score 104.5; DB 2; Length 1822;
Best Local Similarity 24.5%; Pred. No. 98;
Matches 35; Conservative 32; Mismatches 53; Indels 23; Gaps 5;

QY 16 TLNLLGNAPEKALRNEERAIDELKKQAI--EDKEATTATAEASDALEALADQTD-AL 72
Db 1214 SLDTGTGEARDAVELAKDK-----ELAKEATNTEEEATKIVEKLAEDTRKAIDENPNLSD 1269

QY 73 QSEEAHVKADNAASDALEALADQTDALQSEEAHVWQSDNAASDAWEKAATPIALDVKTK 132
Db 1270 EDKQAEIKLTDVAKTKLATIRDNADK-RTQAEAKQA-----LADLEKA 1313

QY 133 KTRKPVVKKEERQNVNLTPTTGE 155
Db 1314 KETQKIADKKAIDRLTLIVKDG 1336
```

Search completed: October 13, 2002, 04:48:06  
Job time : 74.8717 secs

...

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:33:22 ; Search time 22.9572 seconds  
(without alignments)  
268.169 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQPRIIPNGGTLTNL.....KKEERQNVNTLPTTGESNP 159

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Description
1	188	24.2	448 1 SPGL_STRSP
2	187	24.1	593 1 SPGL_STRSP
3	131.5	16.9	309 1 SANT_PLAFN
4	110.5	14.2	774 1 STP_LAMB
5	110	14.2	490 1 CE05_ECOLI
6	110	14.2	490 1 CE10_ECOLI
7	108.5	14.0	421 1 TOLA_ECOLI
8	107	13.8	1679 1 Y109_YEAST
9	105.5	13.6	1120 1 STPR_ECOLI
10	105	13.5	388 1 MRPA_STRPY
11	103	13.3	1189 1 SCII_CHICK
12	102	13.1	929 1 E4L3_MOUSE
13	101.5	13.1	539 1 MY53_HYDAT
14	101.5	13.1	1296 1 ASAL_ENTFA
15	101	13.0	771 1 CALD_CHICK
16	100	12.9	857 1 NFM_CHICK
17	100	12.9	1637 1 MRSP_STAAU
18	99.5	12.8	1337 1 DEXT_STRDO
19	99	12.7	464 1 AK15_RAT
20	99	12.7	896 1 IF2_KLFOX
21	98.5	12.7	407 1 IE68_HSVSA
22	98.5	12.7	897 1 IF2_ENTCL
23	98.5	12.7	1290 1 XCPC_XENLA
24	98	12.6	241 1 NEUM_SERCA
25	96	12.4	1714 1 SYEP_DROME
26	95.5	12.3	708 1 ICAL_HUMAN
27	95.5	12.3	740 1 RNFC_ECOLI
28	95	12.2	198 1 HBHA_MYCTU
29	95	12.2	372 1 TOLA_HAEIN
30	95	12.2	483 1 M6_STRPY
31	95	12.2	564 1 M12_STRPY
32	94.5	12.2	1077 1 HLES_DROME
33	93.5	12.0	166 1 VP06_BPPRD

34	93.5	12.0	516	1	P54_ENTFC
35	93.5	12.0	627	1	ABPX_YEAST
36	93	12.0	576	1	MOES_MOUSE
37	92.5	11.9	361	1	FRB4_SCHPO
38	92.5	11.9	1278	1	DYNA_HUMAN
39	91.5	11.8	763	1	HTR2_HALNI
40	91.5	11.8	764	1	HTR2_HALSA
41	91.5	11.8	2459	1	MAPB_RAT
42	91	11.7	1937	1	MYH8_HUMAN
43	90.5	11.6	516	1	RSP3_CHLRE
44	90.5	11.6	527	1	GI9P_HUMAN
45	90.5	11.6	778	1	HTR6_HALNI

#### ALIGNMENTS

##### RESULT 1

ID	SPGL_STRSP	STANDARD;	PRT;	448 AA.
AC	P06654:			
DT	01-JAN-1988	(Rel. 06, Created)		
DT	01-JAN-1988	(Rel. 06, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Immunoglobulin G binding protein G precursor (IGG binding protein G).			
GN	SPG.			
OS	Streptococcus sp. (Lancefield group G).			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1306;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86304178;	PubMed=3745123;		
RA	Fahnestock S.R., Alexander P., Nagle J., Filpula D.;			
RT	"Gene for an immunoglobulin-binding protein from a group G streptococcus.";			
RT	J. Bacteriol. 167:870-880(1986).			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.07 ANGSTROMS) OF 228-282.			
RX	MEDLINE=94213848;	PubMed=8161530;		
RA	Gallagher T., Alexander P., Bryan P., Gilliland G.L.;			
RT	"Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR.";			
RL	Biochemistry 33:4721-4729(1994).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 293-351.			
RX	MEDLINE=95055731;	PubMed=7966308;		
RA	Derrick J.P., Wigley D.B.;			
RT	"The third IgG-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab.";			
RL	J. Mol. Biol. 243:906-918(1994).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 297-352.			
RX	MEDLINE=95308043;	PubMed=7788293;		
RA	Sauer-Eriksson A.E., Kleywegt G.J., Uhlen M., Jones T.A.;			
RT	"Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG.";			
RL	Structure 3:265-278(1995).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 292-352.			
RA	Butterworth S., Lamzin V.S., Wigley D.B., Derrick J.P., Wilson K.S.;			
RT	Submitted (APR-1997) to the PDB data bank.			
RN	[6]			
RP	STRUCTURE BY NMR OF 298-351.			
RX	MEDLINE=91335209;	PubMed=1871600;		
RA	Gronenborn A.M., Filpula D.R., Essig N.Z., Achari A., Whitlow M., Wingfield P.T., Clore G.M.;			
RT	"A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G.";			
RL	Science 253:657-661(1991).			
CC	-!- FUNCTION: BINDS TO THE CONSTANT FC REGION OF IGG WITH HIGH AFFINITY.			
CC				









```
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z80559; CA86129.1;
CC PIR: S48385; S48385.
CC SGD: S0001411; MLP2.
CC KW Hypothetical protein.
CC SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;
CC -----
Query Match 13.8%; Score 107; DB 1; Length 1679;
Best Local Similarity 26.8%; Pred. No. 7.7;
Matches 44; Conservative 32; Mismatches 52; Indels 36; Gaps 7;
QY 17 LTNLLGNAPEKL-----ALRNEERAIDELKQAIEDKEATTAEIASSDALEALADQTD 71
DB 1261 LTNELSDLGKLSSENANADLENKFNLLKQAEKLDASKKQQAALTNELNELKAIDK 1320
QY 72 LQS-----ERAAVVKAD-----NAASDALEALADQTDALQSEAEVQSDN 112
DB 1321 LQDOLHFENAKVIDLTDLKLAHELQSEDSYRDHEKDYRTLMEETESLK-KELQIFKTAN 1379
QY 113 AASDAWEKAATPIALDVKTKD-----TKPVVKK-EERQNVNT 149
DB 1380 SSSDAPEK-----LKNVMEKEKRIIDERTKEFEKKIQETLNKST 1419
CC -----
RESULT 9
STFR_ECOLI
ID STFR_ECOLI STANDARD; PRT; 1120 AA.
AC P76072; P77560;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Slide tail fiber protein homolog from lambdaoid prophage Rac.
GN STFR OR B1372.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12.
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino T., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saico N.,
RA Sampaio G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takenoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12.
RX MEDLINE=92409576; PubMed=1528877;
RA O'Toole P., Stenberg L., Rissler M., Lindahl G.;
RT "Two major classes in the M protein family in group A streptococci.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8661-8665(1992).
CC -|- FUNCTION: BINDS IGG MOLECULES OF THE IGI, IG2 AND IG4 SUBCLASSES,
CC AND ALSO BINDS FIBRINOGEN.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -|- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF000234; AAC74454.1; ALT_INIT.
CC EMBL: D90774; BAA14966.1;
CC EMBL: D90775; BAA14975.1;
CC HSP: P04002; IWFA.
CC ECGene: EGI3370; stfr.
CC InterPro: IPR004089; Chemotaxis_transducer.
CC KW Hypothetical protein; Fiber protein; Repeat; Complete proteome.
CC SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64;
CC -----
Query Match 13.6%; Score 105.5; DB 1; Length 1120;
Best Local Similarity 31.5%; Pred. No. 6.3;
Matches 41; Conservative 18; Mismatches 58; Indels 13; Gaps 4;
QY 15 GTITNLGNAPK-----KLARNFEAIDELKQAIEDKEATTAEIASSD-----ALEALA 66
DB 84 GTINLFGAMTEDDARPEALRPFELMVEHVARNASAVAQNTAAAKKSADASTSAREANT 143
QY 67 DQTDALQSEAEVVKADNADALEALADQ-TDALQSEAEVQSDNAAADAWKAATPI 125
DB 144 HAADAADSARAASTSGAQASQAASSAGTASTKATEA-----SKSAAAESSKSAAT 199
QY 126 ALDVKTKDT 135
DB 200 SAGAAKTSET 209
CC -----
RESULT 10
MRP4_STRPY
ID MRP4_STRPY STANDARD; PRT; 388 AA.
AC P30141;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Fibrinogen- and Ig-binding protein precursor (MRP protein).
GN MRP4.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AP4 / SEROTYPE M4;
RX MEDLINE=92269579; PubMed=1588817;
RA Stenberg L., O'Toole P., Lindahl G.;
RT "Many group A streptococcal strains express two different
immunoglobulin-binding proteins, encoded by closely linked genes.
Characterization of the proteins expressed by four strains of
different M-type.";
RL Mol. Microbiol. 6:1185-1194(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AP4 / SEROTYPE M4;
RX MEDLINE=92409576; PubMed=1528877;
RA O'Toole P., Stenberg L., Rissler M., Lindahl G.;
RT "Two major classes in the M protein family in group A streptococci.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8661-8665(1992).
CC -|- FUNCTION: BINDS IGG MOLECULES OF THE IGI, IG2 AND IG4 SUBCLASSES,
CC AND ALSO BINDS FIBRINOGEN.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -|- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```





```

RA Wilson R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Isoform 2 (heart-specific) has the complete spectrin-
CC actin-binding (SAB) domain and fully interacts with spectrin and
CC actin.
CC -!- SUBCELLULAR LOCATION: Membrane-associated.
CC -!- ALTERNATIVE PRODUCTS: 7 isoforms; 1/4.1b-brain (shown here),
CC 2/4.1b-heart, 3/4.1b-kidney, 4/4.1b-brain, 5/4.1b-heart, 6/4.1b-
CC kidney and 7; may be produced by alternative splicing. Isoform 7
CC is inferred from the cDNA sequence of ref.2.
CC -!- TISSUE SPECIFICITY: Highest expression in brain, lower in testis,
CC adrenal gland, heart and kidney. Also present in muscle and
CC epithelial cells. Isoform 1 is expressed in brain, isoform 2 is
CC expressed in heart and isoform 3 is mostly expressed in kidney but
CC also in heart and brain. Isoform 6 seems to be most abundant in
CC kidney while isoforms 4 and 5 are predominantly expressed in heart
CC and brain.
CC -!- MISCELLANEOUS: The complete SAB domain is present only in the
CC heart-specific isoforms (2 and 5).
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF152247; RAD38048.1; -.
DR EMBL: AF177146; AAD51365.1; ALT_INIT.
DR MGD: MGI:103008; Epb4.113.
DR InterPro: IPR000299; Band_4.1.
DR Pfam: PF00373; Band_41; 1.
DR PRINTS: PR00935; BAND41.
DR SMART: SM00295; B41; 1.
DR PROSITE: PS00660; BAND_41_1; 1.
DR PROSITE: PS00661; BAND_41_2; 1.
DR PROSITE: PS00662; BAND_41_3; 1.
DR Structural protein; Cytoskeleton; Glycoprotein; Alternative splicing;
KW Membrane.
FT DOMAIN 115 329 BAND 4.1-LIKE.
FT DOMAIN 396 528 HYDROPHILIC.
FT DOMAIN 559 602 SPECTRIN-ACTIN-BINDING.
FT DOMAIN 777 929 CARBOXYL-TERMINAL (CTD).
FT VARSPLIC 455 472 MISSING (IN ISOFORM 7).
FT VARSPLIC 623 663 MISSING (IN ISOFORM 7).
FT VARSPLIC 559 559 D -> NSLKRIKGENYVYKHNLMLD (IN ISOFORM
FT 2 AND ISOFORM 5).
FT VARSPLIC 547 558 MISSING (IN ISOFORM 2, ISOFORM 3, ISOFORM
FT 5, ISOFORM 6 AND ISOFORM 7).
FT VARSPLIC 894 929 ALAQAIKAKEQHPDMSVTKVVVKETITPEDGED -> E
FT (IN ISOFORM 4, ISOFORM 5, ISOFORM 6 AND
FT ISOFORM 7).
FT CONFLICT 6 10 GSDSE -> RIRLR (IN REF. 2).
FT CONFLICT 28 28 Q -> R (IN REF. 2).
FT CONFLICT 288 288 A -> V (IN REF. 2).
FT CONFLICT 306 306 V -> G (IN REF. 2).
SQ SEQUENCE 929 AA; 103337 MW; F4975FF405D44AE CRC64;

Query Match 13.1%; Score 102; DB 1; Length 929;
Best Local Similarity 26.9%; Pred. No. 8.8;
Matches 42; Conservative 34; Mismatches 64; Indels 16; Gaps 9;

QY 14 GGTLTNLGN-APEKALRNEERAIDELKKQAEKATTAIE---AASDALEALAD-Q 68
DB 484 GTSQNLITVTPKKAA--EEERVEEDRRKKAEEATPVTLNRHECKTDSERTDTAAGE 341
QY 69 TDALQS--EEAAVQKADNAASDALEALADQTDALQSEAEV-VQSDNAASDAWEK--AAT 123
DB 542 TSATSDQEDAEIRKAQDLDTQDELKMKHQTWISLKRFTFLSTETALTNEWKRLSYS 601
QY 124 PIALDVKKTKDT---KPVVVKKEERQNVNLTPTTGE 156

```

```

DB 602 PVRLAARQ-EDADWIEPLVPEETKQSSGKGLMDGSE 636
RESULT 13
MYS3_HYDAT
ID MYS3_HYDAT STANDARD; PRT; 539 AA.
AC P39922;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, clone 203 (Fragment).
OS Hydra attenuata (Hydra) (Hydra vulgaris).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_TaxID=6087;
RN [1]
SEQUENCE FROM N.A.
RA Nakano M.Y., Slidwill R.P.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY. STRONGEST,
CC TO OTHER NON MUSCLE MYOSINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L35595; AAA29216.1; -.
DR HSP: P24733; LWDC.
DR InterPro: IPR002928; Myosin_tail.
DR Pfam: PF01576; Myosin_tail; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding.
FT NON_TER 1 1 GLOBULAR HEAD.
FT DOMAIN <1 38 RODLIKE TAIL.
FT DOMAIN 39 >539 COILED COIL (POTENTIAL).
FT NON_TER 539 520
FT NON_TER 539 539
SQ SEQUENCE 539 AA; 62319 MW; 9C5AD5664060939D CRC64;

Query Match 13.1%; Score 101.5; DB 1; Length 539;
Best Local Similarity 22.8%; Pred. No. 5.3;
Matches 37; Conservative 32; Mismatches 70; Indels 23; Gaps 3;

QY 6 EOPRIIPNGGTLNL-----LQNAPEKL-----ALRNEERAIDELKKQ 43
DB 97 EYDLRTIEDKLNLTQVTKDLESSLEALEKLDGSEHSLVLEEKIQAEKIDELTEK 156
QY 44 ATEDEKATTAIEAASDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSE 103
DB 157 TEELQSNISRLTEKQNRDKQIDTLNEDIRKQDETISKNAEKKHVDEELKDRTEQIQAA 216
QY 104 EAEVQSDNAASDAWEKAAKATPIALDVKKTKDTKPVVVKKEERQ 145
DB 217 E-DKCNLNAKTKNKLESSIREIEQDLKKEKDKSKWLEKEKK 257

RESULT 14
ASAL_ENTFA
ID ASAL_ENTFA STANDARD; PRT; 1296 AA.
AC P17953;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Aggregation substance precursor.
GN ASAL.
OS Enterococcus faecalis (Streptococcus faecalis).

```



















RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirska R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) -> AMP +  
CC -!- CATALYTIC ACTIVITY: ATP + L-proline + tRNA(Pro) -> AMP +  
CC -!- CATALYTIC ACTIVITY: L-prolyl-tRNA(Pro).  
CC -!- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS  
CC COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE  
CC MONOSPECIFIC ISOLEUCYL-, LEUCYL-, GLUTAMINYL-, METHIONYL-, LYSYL-,  
CC ARGINYL-, AND ASPARTYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY  
CC PROTEINS, P18, P48 AND P43.  
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-II  
CC AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II  
CC AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -!- SIMILARITY: CONTAINS 6 "WHEP-TRS" DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; M74104; AA28594.1; -;  
DR EMBL; U59923; AAC7469.1; -;  
DR EMBL; AE003745; AAF56211.1; -;  
DR PIR; S18644; S18644.  
DR HSP; P00962; IGTR.  
DR FlyBase; FBgn0005674; Aats-glupro.  
DR InterPro; IPR002106; AA\_trna\_ligase\_II.  
DR InterPro; IPR004046; GST\_C.  
DR InterPro; IPR004154; HGTP\_anticodon.  
DR InterPro; IPR000738; WHEP-TRS.  
DR InterPro; IPR000924; tRNA-synt\_1c.  
DR InterPro; IPR002314; tRNA-synt\_2b.  
DR InterPro; IPR001412; tRNA-synt\_1.  
DR InterPro; IPR002316; tRNA-synt\_pro.  
DR Pfam; PF00043; GST\_C; 1.  
DR Pfam; PF03129; HGTP\_anticodon; 1.  
DR Pfam; PF00749; tRNA-synt\_1c; 1.  
DR Pfam; PF00587; tRNA-synt\_2b; 1.  
DR Pfam; PF00458; WHEP-TRS; 6.  
DR PRINTS; PR00987; TRNASYNTHGLU.  
DR PRINTS; PR01046; TRNASYNTHGLU.  
DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
DR PROSITE; PS00179; AA\_TRNA\_LIGASE\_II\_1; FALSE\_NEG.  
DR PROSITE; PS00339; AA\_TRNA\_LIGASE\_II\_2; FALSE\_NEG.  
DR PROSITE; PS00762; WHEP-TRS; 6.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Multifunctional enzyme; Repeat.  
FT DOMAIN 170 754  
FT DOMAIN 755 800  
FT DOMAIN 827 872  
FT DOMAIN 901 946  
FT DOMAIN 980 1025  
FT DOMAIN 1055 1100  
FT DOMAIN 1129 1173  
FT DOMAIN 1174 1180  
FT DOMAIN 1207 1714  
FT SITE 209 220  
FT SITE 438 442  
FT SITE

FT BINDING 441  
FT CONFLICT 102  
FT CONFLICT 233  
FT CONFLICT 341  
FT CONFLICT 583  
FT CONFLICT 592  
FT CONFLICT 753  
FT CONFLICT 802  
FT CONFLICT 873  
FT CONFLICT 887  
FT CONFLICT 1201  
FT CONFLICT 1461  
FT CONFLICT 1587  
SQ SEQUENCE 1714 AA; 189197 MW; 6FE8C58045B48ABC CRC64;  
Query Match 12.4%; Score 96; DB 1; Length 1714;  
Best Local Similarity 25.2%; Pred. No. 42;  
Matches 39; Conservative 27; Mismatches 57; Indels 32; Gaps 6;  
QY 28 LALRNEERAI-----DELKKOATEDKEATTAIEAASSDALEALADOTDALQSEEAAYV 80  
DB 1080 LALKGEYKLSGKDWTPDAKSEPAVVKKEASPV--SMASPAKDELTOEINA--OGEKVRRA 1136  
QY 81 KADNAASDALEALADOTDALQSEEAQVQSD-----NAASDAWEKAATP 124  
DB 1137 KGNKAKEVIDAEVAKLLALKAKYKEVTGTDFFVAGRGGGGGGSAKKAPKEAQPAPKP 1196  
QY 125 IALDVKKTKDTPVVKKEKERNQVNTLPTTGESNP 159  
DB 1197 VR---KEPADASGAVKKOTRLG---LEATKEDNLP 1225  
RESULT 26  
ID ICAL\_HUMAN STANDARD; PRT; 708 AA.  
AC P20810;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Calpain inhibitor (Calpastatin) (Sperm HS-17 component).  
GN CAST.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91124109; PubMed=2577276;  
RA Asada K., Ishino Y., Shimada M., Shimojo T., Endo M., Kimizuka F.,  
RA Kato I., Maki M., Hatanaka M., Murachi T.;  
RT "cDNA cloning of human calpastatin: sequence homology among human,  
RT pig, and rabbit calpastatins.";  
RL J. Enzym. Inhib. 3:49-56(1989).  
RN [2]  
RP SEQUENCE OF 1-283 FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=Lung;  
RX MEDLINE=92235069; PubMed=1569094;  
RA Lee W.J., Ma H., Takano E., Yang H.Q., Hatanaka M., Maki M.;  
RT "Molecular diversity in amino-terminal domains of human calpastatin  
RT by exon skipping.";  
RL J. Biol. Chem. 267:8437-8442(1992).  
RN [3]  
RP SEQUENCE OF 101-317 FROM N.A.  
RX MEDLINE=90037002; PubMed=2553724;  
RA Maki M., Bagci H., Hamaguchi K., Ueda M., Murachi T., Hatanaka M.;  
RT "Inhibition of calpain by a synthetic oligopeptide corresponding to  
RT an exon of the human calpastatin gene.";  
RL J. Biol. Chem. 264:18866-18869(1989).  
RN [4]  
RP SEQUENCE OF 125-283 FROM N.A.  
RX MEDLINE=90165962; PubMed=2407243;  
RA Uemori T., Shimojo T., Asada K., Asano T., Kimizuka F., Kato I.,  
RA Maki M., Hatanaka M., Murachi T., Hanzawa H., Arata Y.;

RT \*Characterization of a functional domain of human calpastatin.\*;  
 RL Biochem. Biophys. Res. Commun. 166:1485-1493(1990).  
 RN [5]  
 RP SEQUENCE OF 153-708 FROM N.A. (SHORT ISOFORM), AND REVISIONS.  
 RA Wang L.F.; Wei S.G.; Miao S.F.; Liu Q.Y.; Koide S.S.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 523-708 FROM N.A.  
 RN [7]  
 RP SEQUENCE OF 425-708 FROM N.A.  
 RN [8]  
 RP SEQUENCE OF 244-708 FROM N.A.  
 RP TISSUE-Placenta;  
 RA El-Anine M., Talbot G., Despres N., Asselin C., Boire G.,  
 RL Menard H.A.;  
 RN Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP PHOSPHORYLATION.  
 RX MEDLINE-91139699; PubMed-1995645;  
 RA Adachi Y., Ishida-Takahashi A., Takahashi C., Takano E., Murachi T.,  
 RA Hatanaka M.;  
 RT "Phosphorylation and subcellular distribution of calpastatin in human  
 RT hemocytoblastic system cells.";  
 RL J. Biol. Chem. 266:3968-3972(1991).  
 CC -I- FUNCTION: SPECIFIC INHIBITION OF CALPAIN (CALCIUM-DEPENDENT  
 CC CYSTEINE PROTEASE). PLAYS A KEY ROLE IN POSTMORTEM TENDERIZATION  
 CC OF MEAT AND HAVE BEEN HYPOTHESIZED TO BE INVOLVED IN MUSCLE  
 CC PROTEIN DEGRADATION IN LIVING TISSUE.  
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -I- DOMAIN: HAS FOUR INHIBITORY DOMAINS.  
 CC -I- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -I- PTM: PHOSPHORYLATED.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D16217; BAA03747.1; -;  
 DR EMBL; D50827; BAA09438.1; -;  
 DR EMBL; M86258; AAB59398.1; -;  
 DR EMBL; M28230; AAA52066.1; -;  
 DR EMBL; M28227; AAA52066.1; JOINED.  
 DR EMBL; M28228; AAA52066.1; JOINED.  
 DR EMBL; M28229; AAA52066.1; JOINED.  
 DR EMBL; M33328; AAA52296.1; -;  
 DR EMBL; U26724; AAC50136.2; -;  
 DR EMBL; S73329; AAB32311.1; -;  
 DR EMBL; U38525; AAA80684.1; -;  
 DR EMBL; U31345; AAB60371.1; -;  
 DR EMBL; U31346; AAB60372.1; -;  
 DR PIR; A34428; A34428.  
 DR MIM; 114090; -;  
 DR InterPro; IPR001259; Calpain\_inhib.  
 DR Pfam; PF00748; Calpain\_inhib; 4.  
 KW Repeat; Thiol protease inhibitor; Alternative splicing;  
 KW Phosphorylation.

FT DOMAIN 1 169 L.  
 FT REPEAT 170 222 INHIBITORY DOMAIN 1.  
 FT REPEAT 304 356 INHIBITORY DOMAIN 2.  
 FT REPEAT 446 499 INHIBITORY DOMAIN 3.  
 FT REPEAT 583 636 INHIBITORY DOMAIN 4.  
 FT VARSPLIC 212 224 MISSING (IN SHORT ISOFORM).  
 FT VARIANT 592 592 G -> E.  
 FT /FTID=VAR\_005298.  
 FT CONFLICT 467 467 R -> L (IN REF. 5).  
 FT CONFLICT 486 488 VKD -> GKE (IN REF. 5).  
 FT CONFLICT 543 543 V -> L (IN REF. 5).  
 FT CONFLICT 562 562 MISSING (IN REF. 8).  
 SQ SEQUENCE 708 AA; 76484 MW; CACD759C9284E3EA CRC64;  
 Query Match 12.3%; Score 95.5; DB 1; Length 708;  
 Best Local Similarity 27.3%; Pred. No. 17; Mismatches 64; Indels 35; Gaps 7;  
 Matches 45; Conservative 21;  
 QY 20 LLGNAPEKIALRNFEAIDEL-----KQAIIDEKATTAATAEASSDALEALADQTD 70  
 Db 352 LLPEPEKPKRSESELIDELSEDFDRSECKEKPSPKTEESKAAAPVSEAVSRTS 411  
 QY 71 --ALQS--EAAVVKADNAASDALEALADQTDALQSEAEVQSDNAASDAWEKA----- 121  
 Db 412 MCSIQSAPPEATLKG--TVPDDAVEALA---DSLGRKREADPEDGKPYMDKVKAKAEDR 467  
 QY 122 -----ATPIALDVKTKTKDPVVKKEERQNVNTLPTTCEE 156  
 Db 468 EKLGEKEETIPDYRLKEEVKDKGKPLPKESKEQ----LPPMSD 509  
 RESULT 27  
 RNFC\_ECOLI STANDARD; PRT; 740 AA.  
 ID RNFC\_ECOLI  
 AC P77611;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DE 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Electron transport complex protein rnfc.  
 GN RNFC OR B1629.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MGI655;  
 RX MEDLINE-97426617; PubMed-9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1233-1274(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE-97251357; PubMed-9097039;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).  
 CC -I- FUNCTION: May be part of a membrane complex involved in electron  
 CC transport (By similarity).  
 CC -I- COPACTOR: Blnds 2 4PE-4S clusters (Potential).  
 CC -I- SUBUNIT: Composed of at least six subunits; rnfa, rnfb, rnfc,  
 CC rnfd, rnfe and rnfg (By similarity).  
 CC -I- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).



```
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF074390; AAC26052.1; -.
CC EMBL: 277162; CAB00936.1; -.
CC EMBL: AB008951; AAK44716.1; ALT_INIT.
CC TIGR: MT0493; -.
CC -----
CC Tuberculin; RV0475; -.
CC Cell adhesion; Heparin-binding; Hemagglutinin; Glycoprotein;
CC Virulence; Complete proteome.
CC INIT_MET 0
CC DOMAIN 151 193 ALA/LYS-RICH.
CC CONFLICT 120 120 R -> P (IN REF. 1).
CC SEQUENCE 198 AA; 21403 MW; 513760F6F1EB6042 CRC64;
CC -----
Query Match 12.2%; Score 95; DB 1; Length 198;
Best Local Similarity 28.9%; Pred. No. 4.9; Mismatches 19; Indels 24; Gaps 6;
Matches 46; Conservative 19;
QY 3 SPIEQPRIIPNGTLLNGLNAPEKIALNRBERAIDELKKQATIEDKEATTA-----IEAA 57
Db 47 SRVEESR-----ARLTKLQEDLPQLTELREKFTAEELRKAAGYLEAATSRYNELVERG 101
QY 58 SSDALEALADQTDALQSEEAANVKAADNASDALE-----ALADQTDALQSEEAANVWOS 110
Db 102 EA-ALERLSQ-----QSFEVSARAGYDQAVELTQELGTVASQTRAVGERAAKLVG 156
QY 111 DNAASDAWEKAATPI--ALDVKTKTKDKPVYKKEERQNV 147
Db 157 ELPKKAAPAKKAAPAKKAAPAKKAAPAKKAAPAKKAAPAKKAAPAKKAAPAKKA 195
CC -----
RESULT 29
ID TOLA_HAETIN STANDARD; PRT; 372 AA.
AC P44678; P94810;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TOLA protein.
GN TOLA OR H10383
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT *whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.;
RL Science 269:496-512(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1479;
RX MEDLINE=97080550; PubMed=8921895;
RA Sen K., Sikkema D.J., Murphy T.F.;
RT *Isolation and characterization of the Haemophilus influenzae toIQ,
RT toIR, toLA and toLB genes.;
RL Gene 178:75-81(1996).
CC -I- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC (Potential).
CC -----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U32722; AAC22041.1; -.
CC EMBL: U32470; AAC44596.1; -.
CC HSSP: P04002; IATP.
CC TIGR: H10383; -.
CC Transport; protein transport; Transmembrane; Repeat; Inner membrane;
CC Complete proteome.
CC DOMAIN 1 8 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 9 29 POTENTIAL.
CC DOMAIN 30 372 PERIPLASMIC (POTENTIAL).
CC VARIANT 48 48 V -> A (IN STRAIN 1479).
CC VARIANT 142 142 K -> R (IN STRAIN 1479).
CC VARIANT 165 165 A -> P (IN STRAIN 1479).
CC VARIANT 190 190 A -> R (IN STRAIN 1479).
CC VARIANT 203 203 V -> A (IN STRAIN 1479).
CC VARIANT 227 227 D -> A (IN STRAIN 1479).
CC VARIANT 232 232 A -> AKAAAEAKAKA (IN STRAIN 1479).
CC VARIANT 234 234 T -> A (IN STRAIN 1479).
CC VARIANT 249 249 L -> F (IN STRAIN 1479).
CC VARIANT 254 254 I -> V (IN STRAIN 1479).
CC VARIANT 306 306 N -> S (IN STRAIN 1479).
CC VARIANT 323 323 T -> A (IN STRAIN 1479).
CC VARIANT 333 333 S -> P (IN STRAIN 1479).
CC SEQUENCE 372 AA; 39831 MW; 266ECF05C6C95544 CRC64;
CC -----
Query Match 12.2%; Score 95; DB 1; Length 372;
Best Local Similarity 30.9%; Pred. No. 9.5; Mismatches 20; Indels 18; Gaps 5;
Matches 38; Conservative 20;
QY 26 EKIALNRBERAIDELKKQATIEDKEATTAIEAASDALE--ALADQTDALQSEEAANVVKAD 83
Db 106 EELKROEQEQRQEQEIKKQ-----EQARQEALEKQKQAEAKAKQAAEAALKAD 155
QY 84 NAASDALEALADQTDALQSEEAANVWOSDNASDAWEKAATPIALDVK-KTKDTKPVVKKK 142
Db 156 -AEAKRLAAAKQAE---EAKAKAAEIAAQKQAEAKAKLEAEAKAKAEAKAKA 210
QY 143 ERQ 145
Db 211 EAE 213
CC -----
RESULT 30
M6_STRPY STANDARD; PRT; 483 AA.
ID M6_STRPY
AC P08089;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE M protein, serotype 6 precursor.
GN EMM6.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86111835; PubMed=3511046;
RA Hollingshead S.K., Fischetti V.F., Scott J.R.;
RT *Complete nucleotide sequence of type 6 M protein of the group A
RT Streptococcus. Repetitive structure and membrane anchor.;
RL J. Biol. Chem. 261:1677-1686(1986).
RN [2]
RP SEQUENCE OF 43-122 FROM N.A.
RX MEDLINE=85166224; PubMed=3885219;
RA Scott J.R., Pulliam W.M., Hollingshead S.K., Fischetti V.A.;
```

```
RT *Relationship of M protein genes in group A streptococci.*;
RL Proc. Natl. Acad. Sci. U.S.A. 82:1822-1826(1985).
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
CC PHAGOCYTOSIS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER M PROTEINS.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M11338; AAA26920.1; -.
CC PIR: A26297; A26297.
CC InterPro: IPR001899; Gram_pos_anchor.
CC InterPro: IPR003345; M_repeat.
CC Pfam: PF00746; Gram_pos_anchor; 1.
CC Pfam: PF02370; M; 9.
CC PRINTS: PR00015; GPOSANCHOR.
CC PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
CC Virulence: Phagocytosis; Cell wall; Repeat; Antigen; Transmembrane;
CC Coiled coil; Signal.
CC SIGNAL 1 42
CC CHAIN 43 483 M PROTEIN, SEROTYPE 6.
CC DOMAIN 43 457 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 458 477 MEMBRANE ANCHOR.
CC DOMAIN 478 483 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 69 138 10 X 7 AA TANDEM REPEATS.
CC DOMAIN 157 269 4.5 X 25 AA TANDEM REPEATS.
CC DOMAIN 279 347 TWO DIRECTLY REPEATED 27 AMINO ACID
CC BLOCKS SEPARATED BY 15 AMINO ACIDS.
CC DOMAIN 348 411 HYDROPHILIC.
CC DOMAIN 412 448 GLY/PRO-RICH (CELL WALL-SPANNING).
CC DOMAIN 449 454 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
CC PROTEINS.
CC SEQUENCE 483 AA; 53472 MW; 68f87f28db53a448 CRC64;
Query Match 12.2%; Score 95; DB 1; Length 483;
Best Local Similarity 29.6%; Pred. No. 13;
Matches 48; Conservative 22; Mismatches 62; Indels 30; Gaps 9;
QY 21 LGNAPEKALNRNEAIDELKKQATE-----DKATTATAEASSD--ALEAL---A 66
DB 304 LANLTAELDKVKEEKQISDASRQGLRRDLDSAREAKKQVEKALEANSKLALEKLNKEL 363
QY 67 DOTDALQSEEAAYVKAD--NAASDAL--EALADQTDALQSEEAAYVQSDNAASDAWE----- 119
DB 364 EESKRLTEKEAEALQAEAKALEKQAEALAKLAGKA--SDSOTPDAPKPGNKVV 422
QY 120 --KAATPIALDVKKTKDKTPVVKKEERQNVNTLPSTGTEGNSP 159
DB 423 PKGQAPQA--GTPKNQNKAPMKETKQ-----LPSTGETANP 458
RESULT 31
M12_STRPY STANDARD; PRT; 564 AA.
AC P19401.
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE M protein, serotype 12 precursor (Fragment).
GN EM12.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
```

```
OX NCBI_taxid=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS24 / Serotype M12;
RX MEDLINE=88058777; PubMed=2445730;
RA Robbins J.C., Spanier J.G., Jones S.J., Simpson W.J., Cleary P.P.;
RT "Streptococcus pyogenes type 12 M protein gene regulation by upstream
RL J. Bacteriol. 169:5633-5640(1987).
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
CC PHAGOCYTOSIS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER M PROTEINS.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M18269; AA88573.1; -.
CC PIR: A60115; A60115.
CC InterPro: IPR001899; Gram_pos_anchor.
CC InterPro: IPR003345; M_repeat.
CC Pfam: PF00746; Gram_pos_anchor; 1.
CC Pfam: PF02370; M; 9.
CC PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
CC Virulence: Phagocytosis; Cell wall; Repeat; Antigen; Transmembrane;
CC Coiled coil; Signal.
CC SIGNAL 1 41
CC CHAIN 42 >564 M PROTEIN, SEROTYPE 12.
CC DOMAIN 42 550 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 551 >564 MEMBRANE ANCHOR.
CC DOMAIN 44 505 COILED COIL (POTENTIAL).
CC DOMAIN 505 541 GLY/PRO-RICH (CELL WALL-SPANNING).
CC DOMAIN 542 547 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
CC PROTEINS.
CC NON_TER 564 564
CC SEQUENCE 564 AA; 62904 MW; 5f1549daca77b46 CRC64;
Query Match 12.2%; Score 95; DB 1; Length 564;
Best Local Similarity 29.6%; Pred. No. 15;
Matches 48; Conservative 22; Mismatches 62; Indels 30; Gaps 9;
QY 21 LGNAPEKALNRNEAIDELKKQATE-----DKATTATAEASSD--ALEAL---A 66
DB 337 LANLTAELDKVKEEKQISDASRQGLRRDLDSAREAKKQVEKALEANSKLALEKLNKDL 456
QY 67 DOTDALQSEEAAYVKAD--NAASDAL--EALADQTDALQSEEAAYVQSDNAASDAWE----- 119
DB 457 EESKRLTEKEAEALQAEAKALEKQAEALAKLAGKA--SDSOTPDAPKPGNKAV 515
QY 120 --KAATPIALDVKKTKDKTPVVKKEERQNVNTLPSTGTEGNSP 159
DB 516 PKGQAPQA--GTPKNQNKAPMKETKQ-----LPSTGETANP 551
RESULT 32
HLES_DROME STANDARD; PRT; 1077 AA.
AC Q02308;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Hairless protein.
GN H.
OS Drosophila melanogaster (Fruit fly).
```





CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X16421; CAA344442.1; ALT\_INIT.  
 DR PIR; S05542; S05542.  
 DR MEROPS; C40.001; -.  
 DR InterPro; IPR000064; NLPC\_P60.  
 DR Pfam; PF00877; NLPC\_P60; I.  
 KW Signal; Cell wall.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 516 P54 PROTEIN.  
 FT ACT\_SITE 429 429 POTENTIAL.  
 SQ SEQUENCE 516 AA; 54596 MW; 402ECAA439846D26 CRC64;

Query Match 12.0%; Score 93.5; DB 1; Length 516;  
 Best Local Similarity 25.0%; Pred. NO. 17;  
 Matches 41; Conservative 29; Mismatches 79; Indels 15; Gaps 4;  
 QY 2 DSPIEQPTIPNGTILNLGNAPKALRN-----EERAIDELKKQAIEDKATTAI 54  
 DB 167 ENAKOKELADNOALESQKGLLAKQADLVKTSLAEEQATAEDKKRADLNKKRAEAE 226  
 QY 55 EAASSDALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVQSDNAA 114  
 DB 227 EQARIREQARLAQA-----RQAAQEKAKKEAREQAAQAQAL-SSASTTSSSAA 281  
 QY 115 SDAEKAAATPIALDVKKYKDFPVVKKERQNVNLTPTGGSN 158  
 DB 282 QSSSEESKAP-----ESTTEESTSTSTTENSSTGSSSTESS 322

RESULT 35  
 ABPX\_YEAST  
 ID ABPX\_YEAST STANDARD; PRT; 627 AA.  
 AC Q08641; Q08644;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Actin-binding protein ABP140.  
 GN ABP140 OR YOR239W/YOR240W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes.  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE=97127829; PubMed=8972580;  
 RA Boyer J., Michaux G., Fairhead C., Gaillon L., Dujon B.;  
 RT "Sequence and analysis of a 26.9 kb fragment from chromosome XV of  
 RT the yeast *Saccharomyces cerevisiae*.";  
 RL Yeast 12:1575-1586(1996).  
 RN [2]  
 RP PARTIAL SEQUENCE FROM N.A., SEQUENCE OF 1-12; 38-72; 102-115; 244-286;  
 RP 560-574 AND 598-608, FUNCTION, AND SUBCELLULAR LOCATION.  
 RC STRAIN=BJS457;  
 RX MEDLINE=98127445; PubMed=9467951;  
 RA Asakura T., Sasaki T., Nagano F., Satoh A., Obaishi H., Nishio H.,  
 RA Imamura H., Hotta K., Tanaka K., Nakanishi H., Takai Y.;  
 RT "Isolation and characterization of a novel actin filament-binding  
 RT protein from *Saccharomyces cerevisiae*.";  
 RL Oncogene 16:121-130(1998).  
 CC -1- FUNCTION: Binds F-actin and shows weak F-actin crosslinking  
 CC activity.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasm and cortical cytoskeleton.  
 CC -1- MISCELLANEOUS: A RIBOSOMAL FRAMESHIFT OCCURS BETWEEN THE CODONS  
 CC FOR LEU-276 AND GLY-277.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; Z75147; CAA99460.1; ALT\_SEQ.  
 DR EMBL; Z75147; CAA99461.1; ALT\_SEQ.  
 DR SGD; S0005765; ABP140.  
 DR InterPro; IPR001601; Meth-transf.  
 DR InterPro; IPR000051; SAM\_bind.  
 KW Actin-binding; Cytoskeleton; Ribosomal frameshift.  
 FT INIT\_MET 0  
 SQ SEQUENCE 627 AA; 71354 MW; D4E55F9485412F39 CRC64;

Query Match 12.0%; Score 93.5; DB 1; Length 627;  
 Best Local Similarity 28.1%; Pred. NO. 21;  
 Matches 39; Conservative 23; Mismatches 60; Indels 17; Gaps 5;  
 QY 27 KLALNEPRAIDELKKQAIEDKATTAIEAAS-----SDAL-EALADQTDAL----- 72  
 DB 97 KIYSGENADINVDFQYKEMETGCAEVLAASSVESDAIOEGVAETEGIATPKOKENE 156  
 QY 73 OSEEAAYVKADNAASDALEALADQTDALQSEEAAYVQSDNAAADAEKAAATPIALDVK 130  
 DB 157 KNDSEEEESANNAPEAEYSQSEEDADIEQSNKGTENAEANASQA-NDGSTTTTSKN 215  
 QY 131 KTKDTKPVVKKERQNVNT 149  
 DB 216 KKKKKKKKKKKRNGNVNT 234

RESULT 36  
 MOES\_MOUSE  
 ID MOES\_MOUSE STANDARD; PRT; 576 AA.  
 AC P26041;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Moesin (Membrane-organizing extension spike protein).  
 GN MSN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93055012; PubMed=1429901;  
 RA Sato N., Funayama N., Nagafuchi A., Yonemura S., Tsukita S.,  
 RA Tsukita S.;  
 RT "A gene family consisting of ezrin, radixin and moesin. Its specific  
 RT localization at actin filament/plasma membrane association sites.";  
 RL J. Cell Sci. 103:131-143(1992).  
 RN [2]  
 RP SEQUENCE OF 11-576 FROM N.A.  
 RX MEDLINE=92243764; PubMed=1573844;  
 RA Furthmayr H., Lankes W.T., Amleu M.R.;  
 RT "Moesin, a new cytoskeletal protein and constituent of filopodia: its  
 RT role in cellular functions.";  
 RL Kidney Int. 41:665-670(1992).  
 CC -1- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL  
 CC STRUCTURES TO THE PLASMA MEMBRANE.  
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; S47577; AAA11762.1; -.  
 DR EMBL; M86390; AAA39728.1; -.





RL Mol. Biol. Cell 7:1167-1180(1996).  
 RN [4]  
 RP SEQUENCE OF 18-1278 FROM N.A.  
 RX MEDLINE=99023772; PubMed=9805007;  
 RA Tokito M.K., Holzbaur E.L.F.;  
 RT "The genomic structure of DCTNL, a candidate gene for limb-girdle  
 muscular dystrophy";  
 RT Biochim. Biophys. Acta 1442:432-436(1998).  
 CC -!- FUNCTION: REQUIRED FOR THE CYTOPLASMIC DYNEIN-DRIVEN RETROGRADE  
 CC MOVEMENT OF VESICLES AND ORGANELLES ALONG MICROTUBULES. DYNEIN-  
 CC DYNACTIN INTERACTION IS A KEY COMPONENT OF THE MECHANISM OF AXONAL  
 CC TRANSPORT OF VESICLES AND ORGANELLES.  
 CC -!- SUBUNIT: LARGE MACROMOLECULAR COMPLEX OF AT LEAST 10 COMPONENTS.  
 CC P150 (GLUED) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC  
 CC DYNEIN.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; P150 (SHOWN HERE) AND P135;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: BRAIN.  
 CC -!- PTM: PHOSPHORYLATED.  
 CC -!- SIMILARITY: BELONGS THE DYNACTIN 150 KDA SUBUNIT FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF064205; AAD55811.1; JOINED.  
 DR EMBL; AF064203; AAD55811.1; JOINED.  
 DR EMBL; AF064204; AAD55811.1; JOINED.  
 DR EMBL; X98801; CA667333.1; JOINED.  
 DR EMBL; AF086947; AAD03694.1; JOINED.  
 DR EMBL; AF086927; AAD03694.1; JOINED.  
 DR EMBL; AF086928; AAD03694.1; JOINED.  
 DR EMBL; AF086929; AAD03694.1; JOINED.  
 DR EMBL; AF086930; AAD03694.1; JOINED.  
 DR EMBL; AF086931; AAD03694.1; JOINED.  
 DR EMBL; AF086932; AAD03694.1; JOINED.  
 DR EMBL; AF086933; AAD03694.1; JOINED.  
 DR EMBL; AF086934; AAD03694.1; JOINED.  
 DR EMBL; AF086935; AAD03694.1; JOINED.  
 DR EMBL; AF086936; AAD03694.1; JOINED.  
 DR EMBL; AF086937; AAD03694.1; JOINED.  
 DR EMBL; AF086938; AAD03694.1; JOINED.  
 DR EMBL; AF086939; AAD03694.1; JOINED.  
 DR EMBL; AF086940; AAD03694.1; JOINED.  
 DR EMBL; AF086941; AAD03694.1; JOINED.  
 DR EMBL; AF086942; AAD03694.1; JOINED.  
 DR EMBL; AF086943; AAD03694.1; JOINED.  
 DR EMBL; AF086944; AAD03694.1; JOINED.  
 DR EMBL; AF086945; AAD03694.1; JOINED.  
 DR EMBL; AF086946; AAD03694.1; JOINED.  
 DR MIM; 601143; .  
 DR InterPro: IPR000938; CAP-GLY.  
 DR Pfam; PF01302; CAP\_GLY; 1.  
 DR PROSITE; PS00845; CAP\_GLY\_1; 1.  
 DR PROSITE; PS0245; CAP\_GLY\_2; 1.  
 KW Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton;  
 KW Alternative splicing; Phosphorylation.  
 FT DOMAIN 48 90  
 FT DOMAIN 164 191  
 FT DOMAIN 213 547  
 FT DOMAIN 943 1049  
 FT DOMAIN 1182 1211  
 FT VARSPIC 1 131  
 FT MAQSKRVYSRTPSGRMSASAPLRVGRVGVCKGHR  
 FT GTVAYVGNLTATGKRWGVLDKAKRNDGTVQGRKFTCD  
 FT ECHGIFVRSQIQVFEGADTISPTPDSSASKVLKRGTD  
 FT TKTATSKL -> MMRQ (IN ISOFORM P135).  
 FT VARIANT 163 163  
 FT A -> P.

FT CONFLICT 10 10 /FTID=VAR\_001373.  
 FT CONFLICT 132 138 S -> N (IN REF. 2 AND 3).  
 FT CONFLICT 712 712 MISSING (IN REF. 2 AND 3).  
 FT CONFLICT 712 712 D -> V (IN REF. 2 AND 3).  
 SQ SEQUENCE 1278 AA; 141694 MW; 6DCBA5E67856E4BC CRC64;  
 Query Local 11.9%; Score 92.5; DB 1; Length 1278;  
 Best Local Similarity 28.3%; Pred. No. 52;  
 Matches 28; Conservative 21; Mismatches 39; Indels 11; Gaps 2;  
 OY 28 LALNNEERAIDELKKO-AIEDKEATTA-----TEAASSDALEALADDTDALQSE 76  
 DB 943 LKLEDTREYIKELSKLKIGSEANVRLSLLEKLDLSAAKADERIEKVQTRLEETQ 1002  
 OY 77 AAVVKNADNAADALEALADDTDALQSEEAENVVSDNAAS 115  
 DB 1003 ALLRRKKEFEETMDALQADIDQLEAKELQRLNSQS 1041  
 RESULT 39  
 HTR2\_HALN1 STANDARD; PRT; 763 AA.  
 AC Q9HP81;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis  
 DE protein II) (MPP-II).  
 GN HTR2 OR VNG1765G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 OX NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W., Dale H.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.W., Jung K.-H.,  
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR SENSORY  
 CC RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO LIGHT  
 CC CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE005080; AAG19989.1; .  
 DR InterPro: IPR004089; Chemotaxis\_transducer.  
 DR InterPro: IPR003660; HAMP.  
 DR Pfam; PF00672; HAMP; 1.  
 DR Pfam; PF00015; MCPsignal; 1.  
 DR SMART; SM00304; HAMP; 2.  
 DR SMART; SM00283; MA; 1.  
 KW Transducer; Photoreceptor; Transmembrane; Methylation;  
 KW Complete proteome.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 16 36 POTENTIAL.  
 FT DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).

```

Query Match          11.8%; Score 91.5; DB 1; Length 764;
Best Local Similarity 29.1%; Pred. No. 35;
Matches 37; Conservative 20; Mismatches 53; Indels 17; Gaps 5;

QY 37 IDELKKAIQDEKATTTAEAAASSDALEALADQDTALQSEEAAYVKADNAASDALEALADQ 96
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 505 VDDLASAREEVAETVASLADTQQAASAV---DGRQATEDAVETMDVDVADDA-EAAADA 560

QY 97 TDALQSEEAEEVQSDNNAASDANEKA---ATPIALDVKKT---KDTKPVPKKEERQNVNWL 150
      ||| ||| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 561 MDALDSEMDADIGEIVDIADIADQTNMLNALASIEAARTGADGDGFAYVADE-----V 613

QY 151 PTTGEEES 157
      | ||||
Db 614 KTLAEEES 620

```

Search completed: October 13, 2002, 04:46:23  
Job time : 27.9572 secs

**THIS PAGE BLANK (USPTO)**



R.Olsson, A.; Eliasson, M.; Guss, B.; Nilsson, B.; Hellman, U.; Lindberg, M.; Uhlen, M.  
Eur. J. Biochem. 168, 319-324, 1987  
A:Title: Structure and evolution of the repetitive gene encoding streptococcal protein G  
A:Reference number: S00128; MUID:88029445  
A:Accession: S00128  
A:Molecule type: DNA  
A:Residues: 1-593 <OLS>  
A:CROSS-references: EMBL:X06173; NID:g47084; PIDN:CAA29540.1; PID:g47085  
A:Note: the source is designated as Streptococcus G148  
A:Note: part of this sequence, including the amino end of the mature protein, was confir  
R:Sjoebroing, U.; Falkenberg, C.; Nielsen, E.; Akerstrom, B.; Bjorck, L.  
J. Immunol. 140, 1395-1399, 1988  
A:Title: Isolation and characterization of a 14-kDa albumin-binding fragment of streptoc  
A:Reference number: A27604; MUID:88154455  
A:Accession: A27604  
A:Molecule type: protein  
A:Residues: 62-101 <SJO>  
R:Guss, B.; Eliasson, M.; Olsson, A.; Uhlen, M.; Frej, A.K.; Jornvall, H.; Plock, J.I.;  
EMBO J. 5, 1567-1575, 1986  
A:Title: Structure of the IgG-binding regions of streptococcal protein G.  
A:Reference number: A36314; MUID:86300657  
A:Accession: A26314  
A:Molecule type: DNA  
A:Residues: 114-593 <GUS>  
A:CROSS-references: GB:X04015; NID:g47071; PIDN:CAA27638.1; PID:g47072  
C:Function:  
A:Description: it is part of the cell wall structure of group G streptococci and is cova  
C:Superfamily: M5 protein  
C:Keywords: duplication; membrane protein  
F:1-33/Domain: signal sequence status predicted <SIG>  
F:34-593/Product: protein G #status experimental <MAT>  
F:34-116/Domain: alanine-rich <ALA>  
F:117-290/Domain: AB duplication <DUPL>  
F:117-140/Region: A repeat  
F:141-191/Region: B  
F:192-215/Region: A repeat  
F:216-266/Region: B  
F:267-290/Region: A repeat  
F:303-497/Domain: IgG binding <IGB>  
F:303-357/Region: C repeat  
F:358-372/Region: D  
F:373-427/Region: C repeat  
F:428-442/Region: D  
F:443-497/Region: C repeat  
F:498-567/Domain: proline-rich <PRO>  
F:568-593/Domain: carboxyl-terminal <CTD>

Query Match 24.1%; Score 187; DB 2; Length 593;  
Best Local Similarity 38.9%; Pred. No. 2.4e-05;  
Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;  
QY 1 VDSPIEQPRIINGGTLTNLGNAPKALRNEERAIDELKKAIEDKKEATTAEIAAASD 60  
DB 34 VDSPIEDTPIIRNGBELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93  
QY 61 ALEALADQTDALQSEAEVQSDNAA--AASDALEAL--ADQTDALQSEAEVQSDNAA- 114  
DB 94 AWEAAA-AAADAKAKALKEFNKYSDYKKNLTNNAKTVKGVKDLQAVVESAKKAR 152  
QY 115 -SDAWE-----KAATPIALDVKKTK--DTKPVVKKE 142  
DB 153 ISEATDGLSDFLKSOTPAEDTVKSIELAENKVLNRE 189  
RESULT 3  
S67921  
multiple ligand-binding protein 1 precursor - Streptococcus sp. (serotype C20)  
C:Species: Streptococcus sp.  
A:Variety: serotype C20  
C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S67921  
R:Talay, S.R.; Grammel, M.P.; Chhatwal, G.S.  
Biochem. J. 315, 577-582, 1996

A:Title: Structure of a group C streptococcal protein that binds to fibrinogen, album  
A:Reference number: S67921; MUID:96202013  
A:Accession: S67921  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-439 <TAL>  
A:CROSS-references: EMBL:X84989; NID:g1403574; PIDN:CAAS9349.1; PID:g1403575  
A:Note: the authors translated the codon GAG for residue 90 as Ile  
C:Genetics:  
A:Gene: mlb1  
A:Superfamily: M5 protein  
F:1-36/Domain: signal sequence #status predicted <SIG>  
F:37-439/Product: multiple ligand-binding protein 1 #status predicted <MAT>  
Query Match 19.5%; Score 151.5; DB 2; Length 439;  
Best Local Similarity 27.2%; Pred. No. 0.004;  
Matches 65; Conservative 25; Mismatches 46; Indels 103; Gaps 10;  
QY 19 NLIGNA-----PEKIALRNEERAIDELKK-----QAIEDKEATTAEIAA 57  
DB 181 NLIGNAKDQLKLTTEKESLSKDKKALDERNKVLEASRTTRNRDLEAARDKAKATEAELA 240  
QY 58 SSDA-----LEALADQ-----DALQSEEAHVVKAD----- 83  
DB 241 ETNAKVVDKLEEKQVLEASRKRTRNRDLEAARDKAKATEAELAKANELNLQILEASRTTRNR 300  
QY 84 --NAASDALEA-----LADQTDALQSEAEVQSDNAA--AASDAWEKA----- 121  
DB 301 DLEAARDAKKAVDAELAKLKAEEALKEQLAKQAQAEIEKLKESKEKAPETOTPEKPEVP 360  
QY 122 -----ATPIALDVKKTKTKPKV-----VKKEERQNVNLTPTTGESNP 159  
DB 361 GKPSMPWTGLTPATPIAKDRKKT-DVKPAKANKMVPTDVKKDEKK----LPSTGETVNP 414  
RESULT 4  
A44801  
albumin-binding protein precursor - Streptococcus sp. (fragment)  
C:Species: Streptococcus sp.  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: A44801  
R:Sjoebroing, U. 60, 3601-3608, 1992  
A:Title: Isolation and molecular characterization of a novel albumin-binding protein  
A:Reference number: A44801; MUID:92363555  
A:Accession: A44801  
A:Molecule type: DNA; protein  
A:Residues: 1-323 <SJO>  
A:CROSS-references: GB:M95520; NID:g153554; PIDN:AAA26847.1; PID:g153555  
A:Experimental source: group G, strain BG12  
A:Note: sequence extracted from NCBI backbone (NCBI:110938, NCBI:110939)  
F:12-323/Product: albumin-binding protein (fragment) #status experimental <MAT>  
Query Match 17.8%; Score 138.5; DB 2; Length 323;  
Best Local Similarity 33.9%; Pred. No. 0.021;  
Matches 39; Conservative 25; Mismatches 36; Indels 15; Gaps 4;  
QY 42 KQAIEDKEATTAE---AASSD-----ALEALADQTDALQSEAEVQSDNAA--AASDALEAL 93  
DB 34 ESSINKAQLDSVKEIKASVDRDEITELSAQADQIVSAQADNEAITKAEEDSSKAWAA 93  
QY 94 ADQTDALQSEAEVQSDNAA--AASDAWEKAATPIALDVKKTKDTKPVVKKERQNVN 148  
DB 94 ADQANTAKAEADLAKAEKESDAWEKAA---ALDQAK---QAALKEFDYGV 141  
RESULT 5  
YAZQW7  
S-antigen precursor - malaria parasite (Plasmodium falciparum) (strain NF7/Ghana)  
C:Species: Plasmodium falciparum  
C:Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 09-Jun-2000  
C:Accession: B22011  
R:Cowman, A.F.; Saint, R.B.; Coppel, R.L.; Brown, G.V.; Anders, R.F.; Kemp, D.J.



Db 818 EPEDEP 823

RESULT 9  
H90834  
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: H90854  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: H90854  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-437 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA835231.1; PID:g13361273; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 050952  
C:Genetics:  
A:Gene: ECs1808

Query Match 15.4%; Score 119.5; DB 2; Length 437;  
Best Local Similarity 27.2%; Pred. No. 0.56;  
Matches 50; Conservative 21; Mismatches 66; Indels 47; Gaps 5;

QY 15 GTLTNLLGNAP-----KLALRNEERAIDELKKQATKEDKE----- 49  
DB 84 GTLDFLGAMTDDVVRPEALRRFELMVEEAAHAEAKKNAAGEAETSARNAGISASQAPE 143  
QY 50 ----ATTATAEAASSDADALQSEAAVVKADNAASDALEALADOTDQALQSEEA 105  
DB 144 NAANADTSGDASEARONAEAAAKOSEASSSASAAKAKESLOSATDAELSRT 203  
QY 106 EYVQSDNAASDAWEKAATPIALDVKKTKDTPVVKKEER-----QNVNTLPTT-----GE 155  
DB 204 AESAGNAARDATTAA-----EKARESAESQAEQRIAAEAVNRIPTVVGPPGP 255  
QY 156 ESNP 159  
DB 256 KQEP 259

RESULT 10  
B90835  
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: B90835  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: B90835  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-971 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA835073.1; PID:g13361114; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 050952  
C:Genetics:  
A:Gene: ECs1650

Query Match 15.4%; Score 119.5; DB 2; Length 971;  
Best Local Similarity 30.4%; Pred. No. 1.4;  
Matches 45; Conservative 21; Mismatches 69; Indels 13; Gaps 4;

QY 15 GTLTNLLGNAP-----KLALRNEERAIDELKKQATKEDKEATTAIEAASD-----ALEALA 66  
DB 84 GTLDFLGAMTDDVVRPEALRRFELMVEEAVARNASAVAQNTAAAKKSASDASTSAREAT 143

QY 67 DOTDALQSEEAHVVKADNAASDALEALADQ-TDALQSEEAHVQSDNAASDAWEKAATPI 125  
DB 144 HATDAADSARAASTAGQAASSAQSSAGTASTKATEA-----SKSAAAESSKSAAT 199

QY 126 ALDVKKTKDTPVVKKEERQNVNTLPTT 153  
DB 200 SAGAAKTSSTNAAVSQQAASASTATT 227

RESULT 11  
C85693  
probable membrane protein of prophage CP-933X Z1918 [Imported] - Escherichia coli (st  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: C85693  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: C85693  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-973 <STO>  
A:Cross-references: GB:AE005174; NID:g12514847; PIDN:AAG56007.1; GSPDB:GN00145; UMG:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z1918

Query Match 15.4%; Score 119.5; DB 2; Length 973;  
Best Local Similarity 30.4%; Pred. No. 1.4;  
Matches 45; Conservative 21; Mismatches 69; Indels 13; Gaps 4;

QY 15 GTLTNLLGNAP-----KLALRNEERAIDELKKQATKEDKEATTAIEAASD-----ALEALA 66  
DB 86 GTLDFLGAMTDDVVRPEALRRFELMVEEAVARNASAVAQNTAAAKKSASDASTSAREAT 145

QY 67 DOTDALQSEEAHVVKADNAASDALEALADQ-TDALQSEEAHVQSDNAASDAWEKAATPI 125  
DB 146 HATDAADSARAASTAGQAASSAQSSAGTASTKATEA-----SKSAAAESSKSAAT 201

QY 126 ALDVKKTKDTPVVKKEERQNVNTLPTT 153  
DB 202 SAGAAKTSSTNAAVSQQAASASTATT 229

RESULT 12  
T46231  
hypothetical protein T9C5.130 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T46231  
R:Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zilpp, M.; Mewes, H.W.; Lemck  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: Z23026  
A:Accession: T46231  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-166 <RIE>  
A:Cross-references: EMBL:ALJ32964  
A:Experimental source: cultivar Columbia; BAC clone T9C5  
C:Genetics:  
A:Map position: 3  
A:Introns: 32/3  
A:Note: T9C5.130

Query Match 14.9%; Score 116; DB 2; Length 166;  
Best Local Similarity 26.3%; Pred. No. 0.32;  
Matches 41; Conservative 25; Mismatches 62; Indels 28; Gaps 5;

QY 12 PNGGTLTNLLGNAP-----EKLALRNEERAIDELKKQATKEDKEATTAIEAASDALEALA 66  
DB 8 PRESQMONEGSGVPKPKVSENVAKENNTESEKONQTVAEETTTSVEAKETFPVPEPTK 67









DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: G90898  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-271 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA035582.1; PID:q1361625; GSPDB:GND0154  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: ECs2159

Query Match 14.0%; Score 108.5; DB 2; Length 271;  
 Best Local Similarity 29.0%; Pred. No. 1.8;  
 Matches 51; Conservative 22; Mismatches 54; Indels 49; Gaps 8;

OY 15 GTLTNLGNAP-----KLARNERAIDELKKAIEDKATTAIEAASSDALALADQTD 70  
 DB 86 GTLNDPLGAWSEDDVREARRELEWEEAARAEAKKAGRAETSSARNA-----GIS 139  
 OY 71 ALQSEAAVVKADNAASDALEALADQTD-----ALQSEAEVVSQDNASDAWEKA--ATP 124  
 DB 140 ASQAEESA-ANADTSACDASESARQAESAANKQSEEA-----SSSSAQAQAQASQ 194  
 OY 125 IALDVKKTKDT-----KPVKKKERON-----VNTLPT 152  
 DB 195 SAADAEIKKTAESAACNAARMQRPQOKKPGSOOKAHSGREGRIAEEAVNRIPT 250

RESULT 25  
 JY0057  
 Toia protein - *Escherichia coli*

C:Species: *Escherichia coli*  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 02-Feb-2001  
 C:Accession: JY0057; B64810  
 R:Levengood, S.K.; Webster, R.E.  
 J. Bacteriol. 171, 6600-6609, 1989  
 A:Title: Nucleotide sequences of the *toia* and *tolB* genes and localization of their produ  
 A:Reference number: JY0057; MUID:90078104  
 A:Accession: JY0057  
 A:Molecule type: DNA  
 A:Residues: 1-421 <LEV>  
 A:Cross-references: GB:M28232; NID:9148018; PIDN:AAA24663.1; PID:9148019  
 A:Experimental source: strain JY0057  
 A:Note: The authors translated the initiation codon GTG for residue 1 as Val  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of *Escherichia coli* K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: B64810  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-421 <BLAT>  
 A:Cross-references: GB:AE000177; GB:U00096; NID:q1786955; PIDN:AACT3833.1; PID:q1786960;  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Comment: *toia* and *tolB* proteins are necessary for colicins E2, E3, A, and K to reach t  
 C:Genetics:  
 A:Gene: *toia*  
 A:Map position: 17 min  
 A:Start codon: GTG  
 C:Keywords: nucleotide binding; P-loop; transmembrane protein  
 F:14-34/Domain: transmembrane #status predicted <MSS>  
 F:78-301/Domain: helical #status predicted <HSP>  
 F:355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 14.0%; Score 108.5; DB 2; Length 421;  
 Best Local Similarity 32.6%; Pred. No. 2.9;  
 Matches 42; Conservative 25; Mismatches 43; Indels 19; Gaps 6;

OY 26 EKTLARNEERAIDELKKA-IEDKATTAIEAASSDALALADQTDALQSEAAVVKADN 84  
 DB 109 ERLAIOEQKQKQEAQAELKQAEAAKAAADH-AKKA-EADAKAAEEA-----AKK 162

OY 85 AASDA---LEALADQTDALQSEAEVY-----QSDNAASDAWEKAATPIALDVKKT 132  
 DB 163 AADAKKKAEAEAKAAAEQKRAEAAALAKKAAEAEEAAAFARAKKATTEAEAKKAE 222  
 OY 133 KDTKPVVK 141  
 DB 223 AEKKAEEK 231

RESULT 26  
 T10699

hypothetical protein - common buckwheat  
 C:Species: *Fagopyrum esculentum* (common buckwheat)  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
 C:Accession: T10699  
 R:Fujino, K.; Funatsuki, H.; Inada, M.; Shiono, Y.; Kikuta, Y.  
 submitted to the EMBL Data Library, September 1996  
 A:Description: Expression and sequence analysis of cDNAs corresponding to genes diffe  
 A:Reference number: Z17081  
 A:Accession: T10699  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-159 <FUJ>  
 A:Cross-references: EMBL:D87983; NID:d116566; PID:d1022616  
 A:Experimental source: cv. Kitayuki; Immature seed; 14 days after anthesis

Query Match 13.9%; Score 108; DB 2; Length 159;  
 Best Local Similarity 30.1%; Pred. No. 1;  
 Matches 41; Conservative 26; Mismatches 47; Indels 22; Gaps 6;

OY 12 PNCGTTLNLGNAPKIALRNERAIDELKKAIEDKATTAIEAASSDALALADQTD 71  
 DB 8 PN---TTTGQENKSEVAQVEFAAVQVETPAVE--EAPAAVEAPAAVEEVEEAPA 62  
 OY 72 LQSEAAVVKADNAASDALADQTDALQSEAEVVSQDNASDAWEKAATPIALDVKK 131  
 DB 63 AIEAPAAVE-EBAPAAVEEVA-----ETKEVEEVEVKAA--EAPTEAEAEKK 110  
 OY 132 TKDTKPVKKKERONV 147  
 DB 111 VEEV-----EEKQOV 121

RESULT 27  
 A33939

Ec gamma (Igc) receptor II precursor - *Streptococcus* sp. (fragment)  
 C:Species: *Streptococcus* sp.  
 C:Date: 09-Mar-1990 #sequence\_revision 09-Mar-1990 #text\_change 26-Aug-1999  
 C:Accession: A33939  
 R:Heath, D.G.; Cleary, P.P.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 4741-4745, 1989  
 A:Title: Ec-receptor and M-protein genes of group A streptococci are products of gene  
 A:Reference number: A33939; MUID:89282846  
 A:Accession: A33939  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-405 <HEA>  
 A:Cross-references: GB:M22532; NID:q153628; PIDN:AA895296.1; PID:g552003  
 C:Superfamily: M5 protein  
 C:Keywords: Immunoglobulin receptor

Query Match 13.9%; Score 108; DB 2; Length 405;  
 Best Local Similarity 27.5%; Pred. No. 3;  
 Matches 53; Conservative 25; Mismatches 67; Indels 48; Gaps 10;

OY 6 EDPRIIPNGTTLNLGNA-----PEKLARNERAIDEL--KKQAIED 47  
 DB 218 EIAKIQSEATLNLGSAKRELTDLQAKUDTAAEKAKLESQVTLLENILGSAKRELTD 277  
 OY 48 KEATTAIEAASSDALE-----ALADQTDALQSEEA--AVKADNAASDALE-----ALADQ 96  
 DB 278 LQAKLDAAAEKEKLOSOAAALEKOLENTKRELADLQAKLAATNOEKEKLEAEAKALKEO 337



DB 144 SAMADTSAGDALESARQAESAAMAKOSEBASSASAAAKASESSQSAEAELSRKT 203  
QY 114 ASDAMEKANTPIALDYKTKTDKPVKKBER-----QVNTLPT 152  
DB 204 AESAGCAARADATTATKARESAESASQASQRIAAEAENRPT 248

## RESULT 32

G85631

hypothetical protein 21382 [imported] - Escherichia coli (strain O157:H7, substrain EDL5

C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 14-Sep-2001

C:Accession: G85631

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallantha, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206531

A:Accession: G85631

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-375 &lt;STO&gt;

A:Cross-references: GB:AE005174; NID:912514226; PIDN:AG55515.1; GSPDB:GN00145; UMGCP:213

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: 21382

Query Match

Best Local Similarity 13.6%; Score 105.5; DB 2; Length 375;  
Matches 47; Conservative 21; Mismatches 75; Indels 35; Gaps 4;

QY 15 GTLTNLGNAP-----KLALRNEERAIIDELKKAIDKE----- 49

DB 86 GTLNDFLGMSDEDVPEALRRFELWEVARRAEAKNACGAEARSNAGTSASQAE 145

QY 50 ----ATTATGAASDALEALADQDALQSEBAVVKADNNAADALADQDALQSEBA 105

DB 146 SAMADTSAGDALESARQAESAAMAKOSEBASSASAAAKASESSQSAEAELSRKT 205

QY 106 EYVQSDNKAADAMEKANTPIALDYKTKTDKPVKKBERQVNTLPT-----GEENP 159

DB 206 AESAGCAARADATTATKARESAESASQASQRIAAEAENRPTVVGPPGPKGP 261

## RESULT 33

G64887

probable tail fiber protein GP37 - Escherichia coli

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence-revision 17-Sep-1997 #text-change 21-Jul-2000

C:Accession: G64887; T09189

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A61720; MUID:97426617

A:Accession: G64887

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1122 &lt;BLAT&gt;

A:Cross-references: GB:AE000234; GB:U00096; NID:91787633; PIDN:AC74454.1; PID:91787636;

A:Experimental source: strain K-12, substrain MG1655

R:Alba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai, H.  
; Motomura, K.; Nakade, S.; Nakamura, Y.; Nishimoto, H.; Nishio, Y.; Oshima, T.; Saito,  
moto, Y.; Horiiuchi, T.  
DNA Res. 3, 363-377, 1996

A:Title: A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the

A:Reference number: 216603; MUID:97251357

A:Accession: T09189

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 3-1122 &lt;AIB&gt;

A:Cross-references: GB:AE000234; GB:U00096; NID:91787633; PIDN:AC74454.1; PID:91787636;

Query Match 13.6%; Score 105.5; DB 2; Length 1122;  
Best Local Similarity 31.5%; Pred. No. 14;  
Matches 41; Conservative 18; Mismatches 58; Indels 13; Gaps 4;

QY 15 GTLTNLGNAP-----KLALRNEERAIIDELKKAIDKEATTATIEAASSD-----ALEALA 66

DB 86 GTLNDFLGMSDEDVPEALRRFELWEVARRAEAKNACGAEARSNAGTSASQAE 145

QY 67 DQDALQSEBAVVKADNNAADALADQDALQSEBAVVKADNNAADALADQDALQSEBA 125

DB 146 HAADADNNAADTSAGDALESARQAESAAMAKOSEBASSASAAAKASESSQSAEAELSRKT 201

QY 126 ALDYKTKTDK 135

DB 202 SAGCAATSER 211

## RESULT 34

S57834

fcrA protein precursor - Streptococcus pyogenes

C:Species: Streptococcus pyogenes

C:Date: 28-Oct-1995 #sequence-revision 03-Nov-1995 #text-change 26-Aug-1999

C:Accession: S57834

R:Boyle, M.D.P.; Hawltzky, J.; Raeder, R.; Podbielski, A.  
Infect. Immun. 62, 1336-1347, 1994

A:Title: Analysis of genes encoding two unique type IIA immunoglobulin G-binding prot

A:Reference number: S57834; MUID:94178942

A:Accession: S57834

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-387 &lt;BOY&gt;

A:Cross-references: EMBL:X72932; NID:9507128; PIDN:CA51436.1; PID:9507129

C:Superfamily: M5 protein

Query Match

Best Local Similarity 13.5%; Score 105; DB 2; Length 387;  
Matches 53; Conservative 25; Mismatches 67; Indels 48; Gaps 10;

QY 6 EOPRIIPNGSTLNLGNA-----PKLALRNEERAIIDELKKAIDKE 47

DB 180 ETAKLDSANTLENLSAKRELTDLOAKLDTATKAKLESQVTLLENLSAKRELTD 239

QY 48 KEATTAIEAASDALE-----ALADQDALQSEBA-----AVKADNNAADALE-----ALADQ 96

DB 240 LQAKLDRANAEXEKLOSQAALAEKLEATRKELADLOAKLVATNOEKLEAEAKALKRQ 299

QY 97 TDALQSEBAVVKADNNAADAMEKANTPIALDYKTKTDKPVKKBERQVNTLPT----- 148

DB 300 L-AKQAEELAKLKADKASG-----AKP--DTKPGNKVEYPRPSQTRITNKAASMAQTK 350

QY 149 -TUPPTGEE-SNP 159

DB 351 RELPSTGEETNP 363

## RESULT 35

A46173

M6P4 protein - Streptococcus sp. (group A)

C:Species: Streptococcus sp.

C:Date: 21-Sep-1993 #sequence-revision 25-Apr-1997 #text-change 30-May-1997

C:Accession: A46173

R:O'Toole, P.; Stenberg, L.; Rissler, M.; Lindahl, G.  
Proc. Natl. Acad. Sci. U.S.A. 89, 8661-8665, 1992

A:Title: Two major classes in the M protein family in group A streptococci.

A:Reference number: A46173; MUID:92409576

A:Accession: A46173

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-388 &lt;OIT&gt;

A:Note: sequence extracted from NCBI backbone (NCBIN:114063, NCBI:114064)

C:Superfamily: M5 protein

```

Query Match          13.5%; Score 105; DB 2; Length 388;
Best Local Similarity 26.9%; Pred. No. 4.5;
Matches 52; Conservative 26; Mismatches 67; Indels 48; Gaps 10;

QY 6 EOPRIINGCTLTNLGNA-----PEKLALNERAIDEL---KKQAIED 47
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 EIAKIQSEAAFTLEMLGSAKRELTELQKLDPTATAEAKLESQVLTLENLGSAKRELTD 240
QY 48 KEATTAIEAASSDLEA----LADQTDALQSEEA---AVYVADNAASDALE---ALADQ 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 LQALDLAANAEEKELQSOATLEKOLENTKELDLQAKLAIQOEKLEAEAKALKEQ 300
QY 97 TDALQSEAEVYVQSDNADNAAMEKAATPIALDVKRTKTPVYKKEERQNVN----- 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 L-AAQAEELAKLKADKASG-----AQKP---DTKPKKEVPTRESQRTNTNKKAPMAQTK 351
QY 149 -TLPTGEE-SNP 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 352 ROLPSTGEETTPN 364

RESULT 36
S42574
Streptococcal surface protein - Streptococcus dysgalactiae
C:Species: Streptococcus dysgalactiae
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 19-May-2000
C:Accession: S42574
R:Jonsen, H.; Mueller, H.P.
Eur. J. Biochem. 220, 819-826, 1994
A:Title: The type-III Fc receptor from Streptococcus dysgalactiae is also an alpha(2)-ma
A:Reference number: S42574; MUID:94192673
A:Accession: S42574
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-664 <JON>
A:Cross-references: EMBL:Z29666; MID:9470220; PIDN:CAA82764.1; PID:9470221
C:Superfamily: M5 protein

Query Match          13.5%; Score 105; DB 2; Length 664;
Best Local Similarity 23.9%; Pred. No. 8.3;
Matches 54; Conservative 21; Mismatches 75; Indels 76; Gaps 9;

QY 1 VDSPIEOP-----RIIPNGGTLTNLGNAPKRLALRNERAIDELKQAIED----- 47
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 423 IDABELPALTYKLYKGNTFPS---GETTKAV--DAETAEKAKFOANENGYGVGEMSY 477
QY 48 KEATTAIEAASSDLEAALADQTDALQSEEA-VKADNAASDALEALADQTDALQSEEA-E 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 478 DDAKRTFTVTEKPAVIDAPDELTPALTYKIVY---NGKTLKGETTKAVDAETAEKAFK 533
QY 107 VVQSDNADNAW-----EK-----AATPIADV 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 QYAVENGVGVYDDATKFTVTETEMVTEVPGDAFTEPEKEPASIPVLPAPPIAKDD 593
QY 130 KRTKTDTPV-----VKKEERQNVTLPTTGESNP 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 AKKDDTKVDTKKEDAKKPEAKKEBAKKAATLPTTGESNP 639

RESULT 37
C90769
probable tail fiber protein (imported) - Escherichia coli (strain 0157:H7, substrain RIM
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: C90769
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11538796
A:Accession: C90769
A:Status: preliminary
A:Molecule type: DNA

```

```

A:Residues: 1-439 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA834546.1; PID:913360583; GSPDB:GN00154
A:Experimental source: strain 0157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecs1123

Query Match          13.4%; Score 104.5; DB 2; Length 439;
Best Local Similarity 26.4%; Pred. No. 5.6;
Matches 47; Conservative 20; Mismatches 76; Indels 35; Gaps 4;

QY 15 GTLTNLGNAPD-----KLALNERAIDELKQAIEDK----- 49
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 86 GTLNDLCAMSEDDVREPALRFEIWEAAHAEAKKNGAEATSARNAGISASQAE 145
QY 50 ----ATTATIEAASSDLEALADQTDALQSEEA-VKADNAASDALEALADQTDALQSEEA 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 146 SAANADTSAGDASASARPAALSAANAQAQSEASSSSAQAQAQKASSQSAAEELSKRT 205
QY 106 EVVQSDNADNAAMEKAATPIALDVKRTKTPVYKKEERQNVNLTPTT---GESNP 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 206 AESAAGNAARDA--TTAEKARESEASQSAEQSRIAAEEAVNRLPTVYGPGRKGP 261

RESULT 38
S72375
surface exclusion protein precursor - Enterococcus faecalis plasmid pD1
C:Species: Enterococcus faecalis
C>Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 15-Oct-1999
C:Accession: S72375
R:Hirt, H.; Wirth, R.; Muscholl, A.
Mol. Gen. Genet. 252, 640-647, 1996
A:Title: Comparative analysis of 18 sex pheromone plasmids from Enterococcus faecalis
A:Reference number: S72375; MUID:97074879
A:Accession: S72375
A:Molecule type: DNA
A:Residues: 1-843 <HIR>
A:Cross-references: EMBL:X96976; NID:91272652; PIDN:CAA65662.1; PID:e235489; PID:9127
A:Experimental source: strain OG1x
C:Genetics:
A:Gene: sep1
A:Genome: plasmid pD1
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-843/Product: surface exclusion protein #status predicted <MAT>

Query Match          13.4%; Score 104.5; DB 2; Length 843;
Best Local Similarity 28.3%; Pred. No. 12;
Matches 39; Conservative 27; Mismatches 61; Indels 11; Gaps 5;

QY 7 QPRIPNGGTLTNLGNAPKRLALRNERAIDELKQAIEDKQATTAIEAASSDLEALA 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 32 QPK-----TEPNASSEQPTVKAQTGTTEQALTEKQOYIE-KQAIVDQKQVADPAKKEK 84
QY 67 DQTD-ALQSEEA-VKADNAASDALEALADQTDALQSEEA-EVQSDNADNAAMEKAATPI 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 85 DTIDQSVKDOQAVVNDKDALVQSOQAVTDQ-QAVVDPAKKV--DEATPAIEKAKQV 141
QY 126 ALDVAKKTDTKPYAKKEE 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 142 ATDTQAVDDQKVVQEQ 159

RESULT 39
S33441
EF protein - Streptococcus suis
C:Species: Streptococcus suis
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C:Accession: S33441
R:Smith, H.E.; Reek, F.H.; Vecchi, U.; Gielkens, A.L.J.; Smits, M.A.
submitted to the EMBL Data Library, May 1993
A:Description: Repeats in an extracellular protein of wek-pathogenic strains are abse
A:Reference number: S33441
A:Accession: S33441
A:Status: preliminary

```

A: Molecule type: DNA  
A: Residues: 1-1822 <SMI>  
A: Cross-references: EMBL:X71880; NID:g298031; PIDN:CAA50714.1; PID:g298032

Query Match 13.4%; Score 104.5; DB 2; Length 1822;  
Best Local Similarity 24.5%; Pred. No. 28;  
Matches 35; Conservative 32; Mismatches 53; Indels 23; Gaps 5;

OY 16 TLTNLGNAPKELALNBERAIDELKKQAT--EDKEATTAIEAASSDALEALADQTD-AL 72

DB 1214 SLDTGKEARDVAELAKDK----ELAKEARTEEERATKIVEKLAEDYTRKAIEDNPPLSD 1269

OY 73 QSEEAAYKADNAASDALALADQTDALQSEEAAYVQSDNAASDAWEKAATPIALDVKKT 132

DB 1270 EDKQAEIKKLTDAVAKTLATIRDNADK-RQGEKKAQA-----LADLEKA 1313

OY 133 KQTKPVYKKEERQNVNTLPPTGE 155

DB 1314 KETQKTAIDRAAIDRLTLVKDGE 1336

## RESULT 40

G41662

130K surface exclusion protein Sec10 precursor - Enterococcus faecalis plasmid pCF10

C:Species: Enterococcus faecalis

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999

C:Accession: G41662

R:Kuo, S.M.; Olmsted, S.B.; Vikarins, A.S.; Gallo, J.C.; Dunny, G.M.

J: Bacteriol. 173, 7650-7664, 1991

A:Title: Molecular and genetic analysis of a region of plasmid pCF10 containing positive

terococcus faecalis.

A:Reference number: A41662; MUID:92041679

A:Accession: G41662

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <KAO>

A:Cross-references: GB:M64978; NID:g150552; PIDN:AAA65847.1; PID:g150554

C:Genetics:

A:Genome: plasmid

Query Match 13.4%; Score 104; DB 2; Length 891;

Best Local Similarity 26.9%; Pred. No. 14;

Matches 42; Conservative 27; Mismatches 73; Indels 14; Gaps 6;

OY 7 OPRIPNGGTLTNLGNAPKELALNBERAIDELKKQATIEDKREATTATIEAASSDALEALA 66

DB 32 QRTPEPNSST-----EQPTVAKATQTTTEQALTE-KQQQVTEKQAIYDOKQOVADTAKKER 84

OY 67 DQTDALQSEEAAYKADNAASD-ALFALADQTDALQSEEAAYVQSDNAASDAWEKAATPI 125

DB 85 DAIDQSVKDAQAVVDQNKDALDQSQQAVTDQ-QAVVDEAKKV--DEATPSATEKAKEQV 141

OY 126 ALDVAKTKDKTPVYKKEE--RONVNTLPPTGEEEN 158

DB 142 ATDVTGAVDEQKVVDAQQTDVANOQAVVDEKAKETN 177

Search completed: October 13, 2002, 04:49:16  
Job time : 47.3636 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:54:53 : Search time 51.016 Seconds  
(Without alignments)  
765.558 Million cell updates/sec

Title: US-09-847-539A-6  
Perfect score: 777  
Sequence: 1 VDSPIRQPRIPNGTTLTNL.....KKEPQNVNLTPTGESNP 159

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xih  
-O=/cgn2\_1/USPTO.spool/US09847539/runal.10102002\_093106\_5020/app\_query.fasta.1.526  
-DB-Issued\_Patents\_NA -OEM=fastap -SUFFIX=std.rni -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIG=200 -THR.SCORE=PCT -THR.MAX=100 -THR.MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=PCT -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USBR=US09847539\_ECGN1.1.13\_4numat\_10102002\_093106\_5020 -NCPU=6 -TCP=3  
-NO\_XLPHY -NO\_MMAB -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Issued\_Patents\_NA.\*  
2: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114.5	14.7	1555	3 US-08-669-408B-9	Sequence 9, Appli
2	105.5	13.6	989	4 US-09-121-517B-5	Sequence 1, Appli
3	105	13.5	2526	3 US-08-669-408B-1	Sequence 1, Appli
4	104.5	13.4	6744	1 US-08-119-125A-2	Sequence 2, Appli
5	101	13.0	2706	2 US-08-630-822A-61	Sequence 61, Appli
6	101	13.0	2706	2 US-09-005-069-61	Sequence 61, Appli
7	97.5	12.5	4248	3 US-08-678-614-1	Sequence 1, Appli
8	96.5	12.4	1886	4 US-08-936-165A-224	Sequence 224, App
9	96	12.4	824	1 US-08-158-353-1	Sequence 1, Appli
10	94.5	12.2	937	4 US-08-821-872-1	Sequence 1, Appli
11	93	12.0	1332	2 US-08-795-475-5	Sequence 5, Appli
12	92.5	11.9	1419	2 US-08-216-894-9	Sequence 9, Appli

13	92.5	11.9	1419	4 US-09-115-746-9	Sequence 9, Appli
14	92.5	11.9	3279	4 US-08-446-137B-1	Sequence 1, Appli
15	92.5	11.9	4376	1 US-08-119-125A-1	Sequence 1, Appli
16	91.5	11.8	3256	2 US-08-968-751-3	Sequence 3, Appli
17	91	11.7	3622	2 US-08-591-079-7	Sequence 7, Appli
18	91	11.7	5393	2 US-08-591-079-9	Sequence 9, Appli
19	90.5	11.6	1548	2 US-08-762-106-5	Sequence 5, Appli
20	90.5	11.6	1548	4 US-09-320-774-5	Sequence 5, Appli
21	90.5	11.6	1581	2 US-08-762-106-6	Sequence 6, Appli
22	90.5	11.6	1581	2 US-08-762-106-6	Sequence 6, Appli
23	90	11.6	943	4 US-09-320-774-6	Sequence 6, Appli
24	90	11.6	2085	1 US-08-106-981-1	Sequence 1, Appli
25	90	11.6	2085	1 US-08-072-970-1	Sequence 1, Appli
26	90	11.6	2085	1 US-08-465-746-1	Sequence 1, Appli
27	90	11.6	2085	1 US-08-214-164-1	Sequence 1, Appli
28	90	11.6	2085	1 US-08-469-434-1	Sequence 1, Appli
29	90	11.6	2085	1 US-08-214-222-1	Sequence 1, Appli
30	90	11.6	2085	2 US-08-467-852A-1	Sequence 1, Appli
31	90	11.6	2085	2 US-08-468-718-1	Sequence 1, Appli
32	90	11.6	2085	2 US-08-246-636-1	Sequence 1, Appli
33	90	11.6	2085	2 US-08-247-491A-1	Sequence 1, Appli
34	90	11.6	2085	2 US-08-319-795-1	Sequence 1, Appli
35	90	11.6	2085	2 US-08-468-985-1	Sequence 1, Appli
36	90	11.6	2086	3 US-08-312-949-1	Sequence 1, Appli
37	88.5	11.4	2086	3 US-08-446-201-2	Sequence 2, Appli
38	88.5	11.4	1560	1 US-07-813-584A-1	Sequence 1, Appli
39	88.5	11.4	1560	1 US-08-330-515-1	Sequence 1, Appli
40	88.5	11.4	1771	2 US-08-533-669A-7	Sequence 7, Appli
41	88.5	11.4	1771	2 US-08-511-872-1	Sequence 1, Appli
42	88.5	11.4	2241	2 US-08-785-431-3	Sequence 3, Appli
43	88.5	11.4	2241	4 US-09-205-048-3	Sequence 3, Appli
44	88.5	11.4	2367	2 US-08-785-431-1	Sequence 1, Appli
45	88.5	11.4	2367	4 US-09-205-048-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-08-669-408B-9  
Sequence 9, Application US/08669408B  
Patent No. 6100055  
GENERAL INFORMATION:  
APPLICANT: JONSSON, Bengt  
APPLICANT: LINDERBERG, Martin  
APPLICANT: MOELLER, Hans-Peter  
APPLICANT: RANTAMAKI, Liisa K.  
TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669,408B  
FILING DATE: 03-JUL-1996  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/SE94/00826  
FILING DATE: 06-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 9302855-3  
FILING DATE: 06-SEP-1993  
ATTORNEY/AGENT INFORMATION:



```

Sequence 1 Application US/08669408B
Patient No. 6100055
GENERAL INFORMATION:
APPLICANT: GUSS, Bengt
APPLICANT: JONSSON, Hans
APPLICANT: LINDBERG, Martin
APPLICANT: MUELLER, Hans-Peter
APPLICANT: RANTAMAKI, Liisa K.
TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING
TITLE OF INVENTION: PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/669,408B
FILING DATE: 03-JUL-1996
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: WO PCT/SE94/00826
FILING DATE: 06-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9302855-3
FILING DATE: 06-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 61743/102
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2526 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 506..2497
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 506..2497
US-08-669-408B-1
Alignment Scores:
Pred. No.: 0.00638 Length: 2526
Score: 105.00 Matches: 54
Percent Similarity: 33.1% Conservative: 21
Best Local Similarity: 23.8% Mismatches: 75
Query Match: 13.51% Indels: 76
DB: 3 Gaps: 9
US-09-847-539A-6 (1-159) x US-08-669-408B-1 (1-2526)
QY 1 ValAspSerProIleGluInPro-----ArgIleIleProAsnGlyIy 15
Db 1772 ATTAGCAGCAGCTGAACCTACCTCCAGCGCTTGACTACTTACCAACTATATGTTAAAGGTAAAC 1831
QY 16 ThrLeuThrIrsnLeuLeuGlyAsnAlaIleProGluIysLeuAlaIleLeuAlaArgAsnGluIuArg 35
Db 1832 ACTTCTCA-----GGCGAAACAACACTACTAAAGCAGTA-----GAGCGACAACACT 1876

```

```

Oy 36 AlaIleaspGluLeuLysGluAlaIleGluasp----- 47
Db 1877 GCAGAAAAAGCCCTTCAACCAATACGCTAACAAAAAGCGTTTACGGTGAATGCTTTAT 1936
Oy 48 LysGluAlaThrThrPalaIleGluAlaIleSerSerAspAlaLeuGluAlaLeuAlaasp 67
Db 1937 GACGATGCACATAAAACCTTTACAGTTACTGAAAAACGACAGATGATGACGCACCTGAA 1996
Oy 68 GluThrAspAlaLeuGlnSerGluGluAlaIleValValLysAlaAspAlaIleSer 87
Db 1997 TTAAACACGACATTCACCACTACCAAACTGTGTATC-----AATGTAACAA 2044
Oy 88 AspAlaLeuGluAlaLeuAlaIleAspGlnThrAspAlaLeuGlnSerGluGluAla--Glu 106
Db 2045 TTGAAGGCGCAAAACACTACTAAAGCAGTACACGCAGAACTGCAGAAAAAGCCTTCMAA 2104
Oy 107 ValValGlnSerAspAlaIleAlaIleSerAspAlaTrp----- 118
Db 2105 CAATACGCTAACGAAAAACGGTGTATGCTGTTTGACACTTACGATGATGCACCTAACGC 2164
Oy 119 -----GluLys 120
Db 2165 TTTACGTAAGTAAGAAATGTTACTGACAGTCTCTGATGCACCACTGACCAAGAAAG 2224
Oy 121 -----AlaIleThrProIleAlaLeuAspVal 129
Db 2225 CCAGAGCAGATGATCCCTCTGTTCCGTTAACTCCTGCACACTCCAAATGCTAAAGATGAC 2284
Oy 130 LysLysThrLysAspThrLysProVal----- 138
Db 2285 GCTAAGAAAGACGATCTACAGAAAGTCCGACTACTAAGAAAGAACGCTAAAAACCCAGAA 2344
Oy 139 -----ValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThr 153
Db 2345 GCTAAGAAAGAAAGACGCTAAGAAAGACAGAGCTAAGAAAGCTGCAACTCTTCTACACT 2404
Oy 154 GlyGluGluSerAsnPro 159
Db 2405 GCTGAAGCAAGCAACCCCA 2422

RESULT 4
US-08-119-125A-2
; Sequence 2, Application US/08119125A
; Patent No. 5610011
; GENERAL INFORMATION:
; APPLICANT: SMITH, Hilda Elizabeth
; APPLICANT: VECHE, Uri
; TITLE OF INVENTION: DNA Sequences which code for Virulence
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, po
; TITLE OF INVENTION: antibodies derived therefrom and the use thereof for the di
; TITLE OF INVENTION: protection against infection by S. suis in mammals, includi
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Central Diergeneeskundig Instituut
; STREET: Edelhertweg 15
; CITY: PH Leijstad
; STATE:
; COUNTRY: The Netherlands
; ZIP: NL-8219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS V.6.0
; SOFTWARE: Wordperfect v. 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/119,125A
; FILING DATE: 20-SEP-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL92/00054
; FILING DATE: 19-MAR-1992
; APPLICATION NUMBER: NL 9100510
; FILING DATE: 21-MAR-1991

```

ATTORNEY/AGENT INFORMATION:  
NAME: Handal, Anthony H.  
REGISTRATION NUMBER: 26275  
REFERENCE/DOCKET NUMBER: SMITHHE119125  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (203) 838-8589  
TELEFAX: (203) 838-8794  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6744 base pairs  
TYPE: Nucleic acid with corresponding amino acids  
STRANDEDNESS: single stranded  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus suis type II (pathogenic)  
FEATURE:  
OTHER INFORMATION: Extracellular factor related protein (EF\*) gene  
FEATURE:  
NAME/KEY: promoter -35 region  
LOCATION: bp 66 to 71  
FEATURE:  
NAME/KEY: promoter -10 region  
LOCATION: bp 89 to 94  
FEATURE:  
NAME/KEY: promoter -35 region  
LOCATION: bp 153 to 158  
FEATURE:  
NAME/KEY: promoter -10 region  
LOCATION: bp 176 to 181  
FEATURE:  
NAME/KEY: ribosome binding site  
LOCATION: bp 350 to 356  
FEATURE:  
NAME/KEY: signal peptide  
LOCATION: bp 361 to 498  
FEATURE:  
NAME/KEY: start of repetitive units RI-RI1  
LOCATION: bp 2869, 3097, 3292, 3520, 4087, 4381, 4609, 4837,  
LOCATION: 5065, 5293, 5521:  
FEATURE:  
NAME/KEY: start of repetitive Asn-Pro-Asn-Leu sequences  
LOCATION: bp 2932, 3160, 3355, 3583, 4150, 4444, 4672, 4900,  
LOCATION: 5128, 5356, 5584:  
FEATURE:  
NAME/KEY: dyad symmetry regions  
LOCATION: from bp 6554 to 6566 and from bp 6571 to 6583  
FEATURE:  
NAME/KEY: dyad symmetry regions  
LOCATION: from bp 6611 to 6625 and from bp 6631 to 6644  
US-08-119-125A-2  
Alignment Scores:  
Pred. No.: 0.0262 Length: 6744  
Score: 104.50 Matches: 35  
Percent Similarity: 46.858 Conservative: 32  
Best Local Similarity: 24.488 Mismatches: 53  
Query Match: 13.458 Indels: 23  
Gaps: 5  
US-09-847-539A-6 (1-159) x US-08-119-125A-2 (1-6744)  
QY 16 ThleuThrasleuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluArg 35  
DB 4000 ACCTTGACAGACAGTAAGAGCTAGAGATGCAATTGGCTAGCATAA--- 4056  
QY 36 AlaIleAspGluLeuLysGlnAlaIle-----GluAspLysGluAlaThrThra 53  
DB 4057 -----GAATTACTAGTCAAGCAATCCGAACAGAACAGAACAGCTACTAAATA 4107  
QY 54 TleGluAlaIleSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp---AlaLeu 72  
DB 4108 GTAGAAACAACTTCAGAGACATACGCCAAGCTATGAGAGCAATCCAACTTGAGAT 4167

QY 73 GlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAspAlaLeuGluAla 92  
DB 4168 GAAGTAAAGCAGCGGAATTAATAAGCTAGCTGCGGCAAAACTTTGCAACC 4227  
QY 93 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsn 112  
DB 4228 ATTCGTACCAATGCAGATMAG---CGTACGCAGAGAGCAAAAGCTCAAGCC----- 4278  
QY 113 AlaIleSerAspAlaThrPGLuLysAlaAlaThrProIleAlaLeuAspValLysThr 132  
DB 4279 -----CTAGCAGATCTTGAAAAAGCT 4299  
QY 133 LysAspThrLysProValValLysGluGluArgGlnAsnValAsnThrLeuProThr 152  
DB 4300 AAAGAAACAGAGAAATTCGATGAAGCGCATGATGATGACTGACTGCTGTGAAA 4359  
QY 153 ThrGluLys 155  
DB 4360 GATGTGAG 4368  
RESULT 5  
US-08-630-822A-61  
Sequence 61, Application US/08630822A  
Patent No. 5840695  
GENERAL INFORMATION:  
APPLICANT: FRANK, GLENN R.  
APPLICANT: HUNTER, SHIRLEY WU  
APPLICANT: MALLERFELS, LYNDY  
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS  
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,822A  
FILING DATE: 11-Apr-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CONNELL, GARY J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-17-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2706 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5..2706  
US-08-630-822A-61  
Alignment Scores:  
Pred. No.: 0.0201 Length: 2706  
Score: 101.00 Matches: 41  
Percent Similarity: 46.308 Conservative: 34  
Best Local Similarity: 25.318 Mismatches: 65  
Query Match: 13.004 Indels: 22

DB: 2 Gaps: 5  
US-09-847-539A-6 (1-159) x US-08-630-822A-61 (1-2706)  
QY 17 LeuThAsnLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAla 36  
DB 1478 TTAGAAAAACAATTAATGATGTTTACAGAAAAA-----TTGAAAGGTGAATTAGACAC 1531  
QY 37 IleAspGluLeuLysGlnAlaIleGluAspLysGluAlaIleThrAlaIleGluAla 56  
DB 1532 AATCAGAAATTAAGAAGCAAGCTGTGAGCTTGAAGTGTCTGCTACTGAAACA 1591  
QY 57 AlaSerSer-----AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 73  
DB 1592 CTGAATTAATGAATTAACAGAACTATGCAAGGTTTACAAACAAAGAGATGCTTACAA 1651  
QY 74 SerGluGluAlaIleAlaValAla-----LysAlaAspAsnAlaAla 86  
DB 1652 CAAGAAGTAGCATCTCTCCAAAGCAAACTTTCTCAAGACAGAGAGCTTACATCAGAGCT 1711  
QY 87 SerAsp-----AlaLeuGluAlaLeuAlaAspGlnThrAspAla 99  
DB 1712 TCTGATATGCAAGATAGACTAGAACCAAAATTCAGGCTCTCATATTGAACTGAGACAT 1771  
QY 100 LeuGlnSerGluGluAlaGluValAlaGlnSerAspAsnAlaAlaSerAsp----- 116  
DB 1772 GTCAGAAATTTGAGAACCAAGTTACCCAAAGACAGACAACTATTGAAAGATATCA 1831  
QY 117 AlaTrpGluLysAlaIleThrProIleAlaLeuAspValLysLysThrLysAspThrLys 136  
DB 1832 ACATTTGAGAAAGAAATGCTCTCTAGAAATTAAGTGAACACACAAACAAATAT 1891  
QY 137 ProValValLysLysGluGluArgGlnAsnValAlaAsnThrLeuProThrThrGluGlu 156  
DB 1892 GAGCAAGAGCTCAAGACATCGCGAAATCGAAAAATCAAGACTGCTCAATAAGAA 1951  
QY 157 SerAsn 158  
DB 1952 GCAAAAT 1957  
RESULT 6  
US-09-005-069-61  
Sequence 61, Application US/09005069  
Patent No. 5932470  
GENERAL INFORMATION:  
APPLICANT: FRANK, GLENN R.  
APPLICANT: HUNTER, SHIRLEY WU  
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS  
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/005, 069  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/630, 822  
FILING DATE: 11-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: CONNELL, GARY J.  
REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2618-17-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2706 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5..2706  
US-09-005-069-61  
Alignment Scores:  
Pred. No.: 0.0201 Length: 2706  
Score: 101.00 Matches: 41  
Percent Similarity: 46.30% Conservative: 34  
Best Local Similarity: 25.31% Mismatches: 65  
Query Match: 13.00% Indels: 22  
Gaps: 5  
US-09-847-539A-6 (1-159) x US-09-005-069-61 (1-2706)  
QY 17 LeuThAsnLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAla 36  
DB 1478 TTAGAAAAACAATTAATGATGTTTACAGAAAAA-----TTGAAAGGTGAATTAGACAC 1531  
QY 37 IleAspGluLeuLysGlnAlaIleGluAspLysGluAlaIleThrAlaIleGluAla 56  
DB 1532 AATCAGAAATTAAGAAGCAAGCTGTGAGCTTGAAGTGTCTGCTACTGAAACA 1591  
QY 57 AlaSerSer-----AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 73  
DB 1592 CTGAATTAATGAATTAACAGAACTATGCAAGGTTTACAAACAAAGAGATGCTTACAA 1651  
QY 74 SerGluGluAlaIleAlaValAla-----LysAlaAspAsnAlaAla 86  
DB 1652 CAAGAAGTAGCATCTCTCCAAAGCAAACTTTCTCAAGACAGAGAGCTTACATCAGAGCT 1711  
QY 87 SerAsp-----AlaLeuGluAlaLeuAlaAspGlnThrAspAla 99  
DB 1712 TCTGATATGCAAGATAGACTAGAACCAAAATTTGCAAGCTCTCCATATTGAACTGAGACAT 1771  
QY 100 LeuGlnSerGluGluAlaGluValAlaGlnSerAspAsnAlaAlaSerAsp----- 116  
DB 1772 GTCAGAAATTTGAGAACCAAGTTACCCAAAGACAGACAACTATTGAAAGATATCA 1831  
QY 117 AlaTrpGluLysAlaIleThrProIleAlaLeuAspValLysLysThrLysAspThrLys 136  
DB 1832 ACATTTGAGAAAGAAATGCTCTCTAGAAATTAAGTGAACACACAAACAAATAT 1891  
QY 137 ProValValLysLysGluGluArgGlnAsnValAlaAsnThrLeuProThrThrGluGlu 156  
DB 1892 GAGCAAGAGCTCAAGACATCGCGAAATCGAAAAATCAAGACTGCTCAATAAGAA 1951  
QY 157 SerAsn 158  
DB 1952 GCAAAAT 1957  
RESULT 7  
US-08-678-614-1  
Sequence 1, Application US/08678614  
Patent No. 6013507  
GENERAL INFORMATION:  
APPLICANT: Delencastre, Herminia  
APPLICANT: Tomasz, Alexander  
TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF  
TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:



```

Query Match: 12.36% Indels: 30
DB: 1 Gaps: 8

US-09-847-539A-6 (1-159) x US-08-158-353-1 (1-824)

QY 14 GlycylthreuthranLeuLeuGlyAsnAlaProGluLys-----LeuAla 29
   ||| ||||| ::| |||||
Db 300 GGGCTAACTCTTACAGAAATAACTAAAAAATTACAGATCTAACGCACTTCTCCGC 359
QY 30 LeuArgAsnGluLys-----ArgAlaIleAspGluLeuLysGlnAlaIleGlu 46
   ::::: ||| ::||| |||||
Db 360 GTCAAAAGATTGCAACTGCTTGCTGCATATGATGAGAGTTGCTAAGAACTATTGGC 419
QY 47 AsplysGluLanThrThrAlaIleGluAlaIleSerSerAsp-----AlaLeuGluAla 64
   ::::: ||| ::||| |||||
Db 420 AATTGTATACCCCAAAATGGTTTAAATGGCCGCTGCTAAACAAACGATCATGTGTACGC 479
QY 65 -----LeuAlaAspGlnThrAspAlaLeuGlnSerGlu 76
   ::|||::: ||| |||||::: |||
Db 480 GGACCTACCTAATATCAACCTTAATATGCGAAGAAATTGATGATGATGAAATTCACAA 539
QY 77 AlaAlaValValLysAlaAspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGln 96
   ||| ::| ||||| ::|
Db 540 GAATTAAGCAAAAATTAATGAA-----GATGCTAAAAAATGTAACAACACTTT 587
QY 97 ThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaIleSerAsp 116
   ||||| |||||::||| |||||::: |||||
Db 588 ACTATTAACCTAAATAAAAGTACTCATGCGGAACCTGGTATGCGAATGAGACCTCTAGTAT 647
QY 117 AlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLys 136
   ||| |||||::||| ::||| |||||
Db 648 GCTAATGCAAAAGCGCT-----ATTTTAAACAATGCTACTAA 689
QY 137 ProValValLysLysGlu--GluArg-----GlnAsnValAsnThrLeuProThrThr 153
   ::||| |||||::: |||||
Db 690 GATTAAGSGTCTCAAGCGTTGAAGGATTATTGATCATCGTAAAAAACCTTGCAAAAGCA 749
QY 154 GlycylGluSer 157
   ::|||:::
Db 750 GCTCAAGAAACA 761

RESULT 10
US-08-821-872-1
; Sequence 1, Application US/08821872A
; Patent No. 6204004
; GENERAL INFORMATION
; APPLICANT: Kaper, James B
; APPLICANT: Jarvis, Karen
; TITLE OF INVENTION: Immunodiagnostic Test for Enterohemorrhagic Escherichia
; TITLE OF INVENTION: coli Infection
; FILE REFERENCE: Seq. ID NO. 62040044 Ref: 1419JK
; CURRENT APPLICATION NUMBER: US/08/821,872A
; CURRENT FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 937
; TYPE: DNA
; ORGANISM: Escherichia coli
US-08-821-872-1

Alignment Scores:
Pred. No.: 0.0282 Length: 937
Score: 94.50 Matches: 46
Percent Similarity: 42.31% Conservative: 20
Best Local Similarity: 29.49% Mismatches: 67
Query Match: 12.16% Indels: 23
DB: 4 Gaps: 6

US-09-847-539A-6 (1-159) x US-08-821-872-1 (1-937)

QY 14 GlycylthreuthranLeuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGlu 33
   ||||| ::||| |||||

```

```

Db      304  GCCCGATTATTTTCATCAGCATTTGGGAGATCTTAGTCTCTTTTCGACGAATGAACAAACGCG 365
QY      34  GIATGTAIIAIAAPGILUeLlyLSgIlaIaIIeGIuNSpLySGIuaIaIhTThaIa 53
Db      364  GCTAAAGGCGCTGCTCAAGTATTCGAAAAAGCA-----AGCTTCGATCTTCAAAGCT 417
QY      54  IIEGuaIaIaIAseSerAsPaIaLeuGIuIaLeuIaIaSPGInThrAsPaIaLeuGIu 73
Db      418  GCTGCTGGCGCTTCTCAAGCTTGCAAAATRAACCTCTGGTCAAGGTACG----- 465
QY      74  SerGIuGIuIaIaIaIaIaIaIyLSyAsPaSmaIaIaIaSerAsPaIaLeuGIuIaLeu 93
Db      466  -----GAAAGTGTGCTGATGCTGCAGAGAGGAGCATTCAGTGGCATGCAAAAGCATG 519
QY      94  AIAAsPGInThrAsPaIaLeuGIuSerGIuIaIaGIuIaIaIaIaIaIaIaIaIaIa 113
Db      520  GCCACAACAAAGCAAGCA-----GCCAGCGGTGCATCTGGCGTTGCAGATGATGTTGCG 573
QY      114  ---AlaSerAsPaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 125
Db      574  AAACGCTCTGACTTGTGCTGAATAATCTTGACAGACCCGCCGAGAAAGACAAGCAATCAAT 633
QY      126  -----AlaIaIaIaIaIaIyLSyThrLyAsPThrLySProVaIaIyLSy 141
Db      634  AAGTTGTTGAATTCCTGATGAT---AAACTAGCAATACACACAGATTTGTTGCCGTGACC 690
QY      142  GIuGIuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 157
Db      691  AGCTTGCTGCTGAAGGTACGAAACAGCTGCCAACACAACTATCTGAGTCC 738

RESULT 11
US-08-795-475-5
? Sequence 5, Application US/08795475
? Patent No. 5965390
? GENERAL INFORMATION:
? APPLICANT: Bivrock, Lars
? APPLICANT: Sjyrding, Ulf
? TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
? NUMBER OF SEQUENCES: 14
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: SEED AND BERRY LLP
? STREET: 6300 Columbia Center, 701 Fifth Avenue
? CITY: Seattle
? STATE: Washington
? COUNTRY: USA
? ZIP: 98104-7092
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/795,475
? FILING DATE: 11-FEB-1997
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Mcmasters, David D.
? REGISTRATION NUMBER: 33,963
? REFERENCE/DOCKET NUMBER: 100084.402D1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (206) 622-4900
? TELEFAX: (206) 682-6031
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1332 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: unknown
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEICAL: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..1329

```

```

US-08-795-475-5

Alignment Scores:
Pred. No.: 0.0664 Length: 1332
Score: 93.00 Matches: 44
Percent Similarity: 40.34% Conservative: 27
Best local Similarity: 25.00% Mismatches: 47
Query Match: 11.97% Indels: 58
DB: 2 Gaps: 8

US-08-847-539A-6 (1-159) x US-08-795-475-5 (1-1332)

QY 21 LeuGLysnAlaProGLuLysLeuAlaLeuArgAnGLuGLuArgAlaIleAspGLuLeu 40
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 790 TTACCAACACTTGACTGCTGTGAATTTGATTAAAGAAAGAAACAAATTCACACCA 849
QY 41 LysLysGLuAlaIleGLu-----AspLysGLuAlaThr 51
   ::||| ::|
Db 850 AGCCGTCACAGGCTTCGCCCTGACTTGACGCATCACGTCAAGCTAAGAAACAACTTGA 909
   ::||| ::|
QY 52 ThrAlaIleGLuAlaIleSerSerAsp-----AlaLeuGLuAlaLeu-----Ala 66
   |||::||| |||::||| |||::||| |||::||| |||
Db 910 AAACCTTAGAAGAAAGCAACACCAATTAAGCTGCTCTTGAAGAACTTAACAAACAGCTT 969
QY 67 AspIleThrAspAlaLeuGLuInserGLuGLuAlaIleValIleAlaIleAspAsnAlaIle 86
   ::||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 970 GAAACAAAGCAAGAAATTTACAGCAAAAAAGAAAGCTGAACTCAACGCAAACTTGACCA 1029
QY 87 SerAspAlaLeuGLuAlaLeuAlaAspGLuThrAspAlaLeuGLuInserGLuGLuAla 106
   ::|||::||| |||::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1030 GAA-----GCAAAGACACTCAAGAACATTA---GCCAAACACCTGAGAACTTGCA 1088
QY 107 ValValGLuSerAspAsnAlaIleAspAlaThrPGLuLysAlaIleThrProIleAla 126
   ::|||::||| |||::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1081 AAACTAAGACCTGCAAAAGCATCAGAC----- 1107
QY 127 LeuAspValLysLysThrLysAspThrLysPro----- 137
Db 1108 -----TCACAAACCCCTGATTACAAAACGAAACAAAGCTCTTCACAGTAAAGT 1158
QY 138 -----ValValLysLysGLuGLu 143
Db 1159 CAGCAGCCACCAAGCAGGTACAAAACCTAACCAAAACAAAGCAACATGAGAAACCTAAG 1218
QY 144 ArgGLuAsnValAsnThrLeuProThrThrGLuGLuSerAsnPro 159
   ||||| |||||::||| |||||::||| |||||::||| |||||::||| |||||::||| |||||::|||
Db 1219 AGACAG-----TTACCATCAACAGGTGAAGACGTAAACCA 1254

RESULT 12
US-08-216-894-9
: Sequence 9, Application US/08216894
: Patent No. 5876734
: GENERAL INFORMATION:
: APPLICANT: Kirchhoff, Louis V.
: APPLICANT: Otsu, Kelo
: TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
: TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington, D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/216,894
: FILING DATE: 24-MAR-1994
: ATTORNEY/AGENT INFORMATION:

```



```

; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1419 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1416
; US-08-216-894-9

Alignment Scores:
Pred. No.: 0.0823 Length: 1419
Score: 92.50 Matches: 34
Percent Similarity: 45.75% Conservative: 36
Best Local Similarity: 22.22% Mismatches: 72
Query Match: 11.90% Indels: 11
DB: 2 Gaps: 4

US-09-847-539a-6 (1-159) x US-08-216-894-9 (1-1419)
QY 3 SerProilegInProArGileleProAsnGlyThrLeuThrAsnLeuGly 22
   ||| : : : ||| ||||| ||| : : :
Db 655 TCGGATGTCGATCGAAGTCTCGGGATCCCCCGGCTGCGAGATTCACAGAGTTGCC 714
QY 23 AsnAlaProGlu-----LysLeuAlaLeuArgAsnGluGluArg 35
   ||| : : : ||| ||||| ||| : : :
Db 715 GAACCGGAGAGCAGAGAGCAGCTGAAGCCAGAGGTTGCCGAGGAGAGACAGAGG 774
QY 36 AlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGlu 55
   ||| : : : ||| ||| : : : ||| ||| : : :
Db 775 GCACCTGAAGCCAGAGAGTTCGCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 834
QY 56 AlaIleAspSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 75
   ||| : : : ||| ||| : : : ||| ||| : : :
Db 835 GTTCCGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891
QY 76 GluAlaIleAlaValIleLysAlaAspAlaIleAspAlaLeuGluAlaLeuAlaAsp 95
   : : : : : ||| : : : ||| ||| : : : ||| : : :
Db 892 AAGCAAAAGCAGCTGAGAGCCAGAGAGTTCGCCGAGAGAGAGAGAGAGAGAGAG 951
QY 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAlaIleAsp 115
   ||| : : : ||| : : : ||| : : : ||| : : :
Db 952 GCCACGAGAGTTCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011
QY 116 AspAlaIleProGluLysAla-----AlaThrProIleAla--LeuAspValLysLysThr 132
   : : : ||| ||| : : : ||| ||| : : : ||| : : :
Db 1012 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071
QY 133 LysAspThrLysProValValLysLysGluGluArgGln 145
   : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1072 GCTGAGAGCAGAGAGAGTTCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1110

RESULT 13
US-09-115-746-9
; Sequence 9, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchoff, Louis V.
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSTIC INFECTION
; TITLE OF INVENTION: WITH TRIPANOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner

```

```

; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1419 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1416
; US-09-115-746-9

Alignment Scores:
Pred. No.: 0.0823 Length: 1419
Score: 92.50 Matches: 34
Percent Similarity: 45.75% Conservative: 36
Best Local Similarity: 22.22% Mismatches: 72
Query Match: 11.90% Indels: 11
DB: 4 Gaps: 4

US-09-847-539a-6 (1-159) x US-09-115-746-9 (1-1419)
QY 3 SerProilegInProArGileleProAsnGlyThrLeuThrAsnLeuGly 22
   ||| : : : ||| ||||| ||| : : :
Db 655 TCGGATGTCGATCGAAGTCTCGGGATCCCCCGGCTGCGAGATTCACAGAGTTGCC 714
QY 23 AsnAlaProGlu-----LysLeuAlaLeuArgAsnGluGluArg 35
   ||| : : : ||| ||||| ||| : : :
Db 715 GAACCGGAGAGCAGAGAGCAGCTGAAGCCAGAGGTTGCCGAGGAGAGACAGAGG 774
QY 36 AlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGlu 55
   ||| : : : ||| ||| : : : ||| ||| : : :
Db 775 GCACCTGAAGCCAGAGAGTTCGCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 834
QY 56 AlaIleAspSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 75
   ||| : : : ||| ||| : : : ||| ||| : : :
Db 835 GTTCCGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891
QY 76 GluAlaIleAlaValIleLysAlaAspAlaIleAspAlaLeuGluAlaLeuAlaAsp 95
   : : : : : ||| : : : ||| ||| : : : ||| : : :
Db 892 AAGCAAAAGCAGCTGAGAGCCAGAGAGTTCGCCGAGAGAGAGAGAGAGAGAGAG 951
QY 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAlaIleAsp 115
   ||| : : : ||| : : : ||| : : : ||| : : :
Db 952 GCCACGAGAGTTCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011
QY 116 AspAlaIleProGluLysAla-----AlaThrProIleAla--LeuAspValLysLysThr 132
   : : : ||| ||| : : : ||| ||| : : : ||| : : :
Db 1012 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071

```

```

QY 133 LysAspThrLysProValIValLysLysGluGluArgGln 145
DB 1072 GCTGAAGCCAGCAGAGGTTCGCCGAGCGAGAGCAGAGCAG 1110

RESULT 14
US-08-446-137B-1
Sequence 1, Application US/08446137B
Patent No. 6162903
GENERAL INFORMATION:
APPLICANT: Trowern, Angus R.
APPLICANT: Atkinson, Anthony
APPLICANT: Murphy, Jonathan P.
APPLICANT: Laurence, Oliver S.
APPLICANT: Dugleby, Clive J.
TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
FROM L PROTEIN AND THEIR USES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,137B
FILING DATE: 22-MAY-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Mcmasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 100084.406
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Peptostreptococcus asaccharolyticus
STRAIN: 1018
FEATURE:
NAME/KEY: exon
LOCATION: 103..3186
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 280
OTHER INFORMATION: /product= "mature protein L"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: CDS
LOCATION: 103..3186
OTHER INFORMATION: /codon_start= 103
OTHER INFORMATION: /product= "Immature protein L"
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 280..3183
OTHER INFORMATION: /codon_start= 280
OTHER INFORMATION: /product= "mature protein L"
FEATURE:
NAME/KEY: misc_signal

```

```

; LOCATION: 208..279
US-08-446-137B-1
Alignment Scores:
Pred. No.: 0.245
Score: 92.50
Percent Similarity: 41.40%
Best Local Similarity: 26.75%
Query Match: 11.90%
DB: 4 Gaps: 6

US-09-847-539A-6 (1-159) x US-08-446-137B-1 (1-3279)
QY 26 GluLysLeuAlaLeuArgGln-----GluArgAlaIleAspLysLeuLysGln 43
DB 778 GAAAAATTGACGACGACCAAAAGCAAGCAAGCAAGCAATATATCAATTTAAACGCTTA 837
QY 44 AlaIleGluAspLysGlu-----AlaThrThrAlaIleGluAla 56
DB 838 GCAGACAAAACAAAAGAAATTAGCTAGAGAGATATGATTGACGCTACTATTGATGCA 897
QY 57 -----AlaSerSerAspAlaLeuGluAlaLeuAlaAspLThrAspAla 71
DB 898 ATCAATGATATCGTACGTACGACGATGTAATGGAAGAAAGTATCTGAAAAAGAAACA 957
QY 72 LeuGlnSerGluGluAlaAlaValValLysAlaAspAspAlaSerAspAlaLeuGlu 91
DB 958 CCAGAACCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1011
QY 92 AlaLeuAlaAspLThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAsp 111
DB 1012 -----AGCCACACAAAATCCAGAAATTCGCAAAACAAACCAATGATGATGCT 1065
QY 112 AsnAlaAlaSerAspAlaThrPglLysAlaIleThrProIleAlaLeuAspVal----- 129
DB 1066 TACGCTTACGCGAGATGCTTTAAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1125
QY 130 -----LysLysThrLysAspThrLysProVal 138
DB 1126 AAGCGTTAACTTTAAATATTTAAATTCGCTGTAAGAAAGCAAGCAAGCAAGCAAGCA 1182
QY 139 ValLysLysGluGluArgLysAsnValAsnThrLeuProThrThrGlyGlu 155
DB 1183 ---AAGAGAAAGTTAACAATCAAGTTAACTTTGAGATGGAAG 1230

RESULT 15
US-08-119-125A-1
Sequence 1, Application US/08119125A
Patent No. 5610011
GENERAL INFORMATION:
APPLICANT: SMITH, Hilida Elizabeth
APPLICANT: VECCHI, Uri
TITLE OF INVENTION: DNA Sequences which code for Virulence
Characteristics of Streptococcus suis and parts thereof, po
TITLE OF INVENTION: antibodies derived therefrom and the use thereof for the di
TITLE OF INVENTION: protection against infection by S. suis in mammals, includi
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESS: Central Diergeneeskundig Instituut
STREET: Edelhertweg 15
CITY: PH Delftstad
STATE:
COUNTRY: The Netherlands
ZIP: NL-8219
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS v. 6.0
SOFTWARE: Wordperfect v. 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/119,125A
FILING DATE: 20-SEP-1993
CLASSIFICATION: 514

```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/NL92/00054  
FILING DATE: 19-MAR-1992  
APPLICATION NUMBER: NL 9100510  
FILING DATE: 21-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Handal, Anthony H.  
REGISTRATION NUMBER: 26275  
REFERENCE/DOCKET NUMBER: SMITHHE119125  
TELEPHONE: (203) 838-8589  
TELEFAX: (203) 838-8794  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4376 base pairs  
TYPE: Nucleic acid with corresponding amino acids  
STRANDEDNESS: single stranded  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus suis type II (pathogenic)  
FEATURE:  
OTHER INFORMATION: Extracellular protein factor (EF) gene  
NAME/KEY: promoter -35 region  
LOCATION: bp 66 to 71  
FEATURE:  
NAME/KEY: promoter -10 region  
LOCATION: bp 89 to 94  
FEATURE:  
NAME/KEY: promoter -35 region  
LOCATION: bp 153 to 158  
FEATURE:  
NAME/KEY: promoter -10 region  
LOCATION: bp 176 to 181  
FEATURE:  
NAME/KEY: ribosome binding site  
LOCATION: bp 350 to 356  
FEATURE:  
NAME/KEY: signal peptide  
LOCATION: bp 361 to 498  
FEATURE:  
NAME/KEY: mature peptide  
LOCATION: bp 499 to 2890  
FEATURE:  
NAME/KEY: dyad symmetry regions  
LOCATION: from bp 4186 to 4198 and from bp 4203 to 4215  
FEATURE:  
NAME/KEY: dyad symmetry regions  
LOCATION: from bp 4243 to 4257 and from bp 4263 to 4276  
US-08-119-125A-1  
Alignment Scores:  
Pred. No.: 0.357 Length: 4376  
Score: 92.50 Matches: 38  
Percent Similarity: 45.45% Conservative: 27  
Best Local Similarity: 26.57% Mismatches: 45  
Query Match: 11.90% Indels: 34 Gaps: 6  
US-09-847-539A-6 (1-159) x US-08-119-125A-1 (1-4376)  
QY 34 GUAAGAlaIleaspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThraAla 53  
Db 2842 GAAAGGCTATTACAGCTCAAGCG-CAATCGGCGAGGACAA-----GCG 2888  
QY 54 IlleGluAlaIleAspSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 73  
Db 2889 GTAGCTCAATCGCCCAAGATGCTTTCACCA-----GCGAACCA 2930  
QY 74 SerGluGluAlaIleAlaValIleAlaAspAlaIleAspAlaLeuGluAlaLeu 93  
Db 2931 GATGCTAAGATTAAGATTGCTAAGAAATCGACGCTGCTAAGTCAAGCATTCAGCCGAAAT 2990

QY 94 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAsp---Asn 112  
Db 2991 CCAAACTTGACAGATGCGAGAAAGAAATCAAGTCAAGAACCGTACATGCAATGCTAA 3050  
QY 113 AlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle----- 125  
Db 3051 GCTCGACAGATGCAATGATGCTTCAACAAGTCACTCGAAGCGCAATCGCAGACGAC 3110  
QY 126 -----AlaLeuAspValLysLysThrLysAspThrLys 136  
Db 3111 AAAGCGCTAGGCGCCATCGCCAAAGACATTCGTGATCCGGAAA---CAAGATGCTAAG 3167  
QY 137 ProValIlyLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGlu 156  
Db 3168 AACAGATTCCTTAAGAGGAGAAATCCGCTAAGTCAATCT-----GAC 3212  
QY 157 SerAsnPro 159  
Db 3213 TCCAAATCCG 3221  
RESULT 16  
US-08-968-751-3  
Sequence 3, Application US/08968751  
Patent No. 5948643  
GENERAL INFORMATION:  
APPLICANT: Rubinfield, Bonnie  
APPLICANT: Polakis, Paul G.  
APPLICANT: Ligenfelder, Carol  
APPLICANT: Vuong, Terilyn T.  
TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ONYX Pharmaceuticals, Inc.  
STREET: 3031 Research Drive  
CITY: Richmond  
STATE: CA  
COUNTRY: USA  
ZIP: 94806  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,751  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Glotta, Gregory  
REGISTRATION NUMBER: 32,028  
REFERENCE/DOCKET NUMBER: ONYX1024 CG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 262-8710  
TELEFAX: (510) 222-9758  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3256 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 34..2541  
US-08-968-751-3  
Alignment Scores:  
Pred. No.: 0.316 Length: 3256  
Score: 91.50 Matches: 37  
Percent Similarity: 42.14% Conservative: 30  
Best Local Similarity: 23.27% Mismatches: 73  
Query Match: 11.78% Indels: 19

```

DB: 2 Gaps: 4
US-09-847-539a-6 (1-159) x US-08-968-751-3 (1-3256)
QY 16 ThrLeuThrAsnLeuLeuGlyAsnAla-----ProGluLysLeuAlaLeuArgAsn 32
DB 919 ACATGGACGAGCTTGTGCTGCGCTCCGACAACTCCAGAGAACTCCAGCTCAAGAGAC 978
QY 33 GluGluArgAlaIleAspGluLeuLysLysGluAlaIleGluAspLysGluAlaThrThr 52
DB 979 CACGAA---GCCCTGCAGCA---CGTGATGTGGCAATTGAGCAAAAGCAGAGGTTTCT 1032
QY 53 AlaIleGluAlaIleAspSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 72
DB 1033 AGCGTCTCGAACAACTCTCTGCGCCAGTTAGAGAGACTGCCAAAGCCAAAGAACAC 1092
QY 73 GlnSerGluGluAlaIleAlaValLysAlaIleAspAspAlaIleAspAlaLeuGluAla 92
DB 1093 GAGTTGGAAACAGCTCTAGCAACAGATCTCCGGGCTGAGTTGCAATCTTGCCCAAC 1152
QY 93 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu----- 104
DB 1153 ATGCAAGCCACTAAAGAGCTACAGACTACACATACCATTTGTCGCCAGACCTGGCT 1212
QY 105 -----AlaGluValValGlnSerAspAsnAlaIleAspAlaTrp 118
DB 1213 ATGCAAGATGAGTTATTCTGCACTTACCAGACCAATGAGAGAGAGCTCTCAATGC 1272
QY 119 GluLysAlaIleThrProIleAlaLeuAspValLysLysThrLysAspThrLysProVal 138
DB 1273 ValLysAlaGAGATGGCACTAAACATGACAGGAGAACTCCAGCAGCAACGCTGTC 1332
QY 139 ValLysLysGluGluAlaArgGlnAsnValAsnThrLysProThrGlyLysGlnSer 157
DB 1333 CTGGCCAAAGAGGTGCGGACCTGAAAGACCTTGAGATTGTCAGACAGAGAGAT 1389

RESULT 17
US-08-591-079-7
; Sequence 7, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: Zychlinsky, Arturo
; TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,079
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Livanic, Shmuel
; REGISTRATION NUMBER: 33,949
; REFERENCE/DOCKET NUMBER: 15661-20017.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0764
; TELEX: 90-4030 MRSNFOERSM
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3622 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Salmonella typhimurium
; STRAIN: s1dB
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 575..2356
; OTHER INFORMATION: /product="Sips"
; OTHER INFORMATION: /gene="sips"
US-08-591-079-7

Alignment Scores:
Pred. No.: 0.415 Length: 3622
Score: 91.00 Matches: 41
Percent Similarity: 42.66% Conservative: 20
Best Local Similarity: 28.67% Mismatches: 70
Query Match: 11.71% Indels: 12
DB: 2 Gaps: 5

US-09-847-539a-6 (1-159) x US-08-591-079-7 (1-3622)
QY 15 GlyThrLeuThrAsnLeuLeuGlyAsnAla-----ProGluLysLeuAla 29
DB 842 GCCAGGTATGATGACCTTACTGGCGGTGCTCTCACTGAGTCTGCTGCGG 901
QY 30 LeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGluAlaIleGlu---AspLys 48
DB 902 GTA-----TGCAGCGCGATGATTGATGATCAAAAAGAGATGGCTTACGATTCGAAA 955
QY 49 GluAlaThrThrAlaIleGluAlaIle-----SerSerAspAlaLeuGluAlaLeuAla 66
DB 956 GAATTCACAGCGCTCTGGAGAGGCTCAGAGCGCAGCATCTGATGACCACTATTC 1015
QY 67 AspGlnThrAspAlaLeuGlnSerGluGluAlaIleAlaValLysAlaIleAspAsnAla 86
DB 1016 AAAAGACGATACCCGCCAGGTGTTATGACGTCGACCAAAAACGACGACGAG 1075
QY 87 SerAspAlaLeuGluAlaLeuAlaIleAspGlnThrAspAlaLeuGlnSerGluGluAla 106
DB 1076 CAAATTAATTGCAATCGCTGGACCGGCTGACCCGCTATGACCAAGCTGAACCCGG 1135
QY 107 ValValGlnSerAspAsnAlaIleAspAlaTrpGluLysAlaIleThrProIleAla 126
DB 1136 GTAGAAACAGCGCGAAAGAGACGACAGAGCGG-----AAAGAGCCTTATGATTAAGGCC 1189
QY 127 LeuAspValLysLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsn 146
DB 1190 ACGGATGCGACGCTTAACGACGACAGCGCCAAAGCGAAGCCGAGAAAGCGGATAC 1249
QY 147 ValAsnThr 149
DB 1250 ATTCTGACC 1258

RESULT 18
US-08-591-079-9
; Sequence 9, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: Zychlinsky, Arturo
; TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



QY 100 LeuGlnSerGluGluAlaGluValAlaGlnSerAspAsnAlaLaserAspAlaTrpGlu 119  
DB 1246 CTGCTGGAGAGACTGTGGCGCC-----ACGGCGCGCGCGAGCGGAGAG 1290  
QY 120 LysAlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProVal--- 138  
DB 1291 CCGCGCGAGAGACCGCGCGCGCGCGCTGCTGCGCGATGGCGGTGGAGCTGTGGAC 1350  
QY 139 ValLysLysGluGluAlaGlnAsnValAsnThrLeuPro 151  
DB 1351 GTGGAGGCTGAGGCTGGCCCAAGCGGTGGAGGCTGTGCC 1389  
RESULT 20  
US-09-320-774-5  
Sequence 5, Application US/09320774  
Patent No. 6265545  
GENERAL INFORMATION:  
APPLICANT: Jarvik, Jonathan W.  
TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
TITLE OF INVENTION: TAGGING  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Harris Brotman  
STREET: 202 Coast Blvd., Suite 111  
CITY: La Jolla  
STATE: California  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/320,774  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/762,106  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brotman, Harris F.  
REGISTRATION NUMBER: 35,461  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 654-2428  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1548 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-320-774-5  
Alignment Scores:  
Pred. No.: 0.156 Length: 1548  
Score: 90.30 Matches: 47  
Percent Similarity: 39.88% Conservative: 22  
Best Local Similarity: 27.17% Mismatches: 41  
Query Match: 11.65% Indels: 63  
DB: 4 Gaps: 8  
US-09-847-539a-6 (1-159) x US-09-320-774-5 (1-1548)  
QY 35 ArgAlaIleAspGluLeuLysLysGlnAla----- 44  
DB 892 CCGCTGCTGCAGAACGCTGTGGAGCGCGCGCGCGCTGCGAGCAATCGCAGCACC 951  
QY 45 IleGluAspLysGluAlaThrAla----- 53

DB 952 CTGGCGGACAAAGCCGCCAGCACGGCGCCAGCTGAGACCGCTGGCGGAGCGGAGGCC 1011  
QY 54 -----IleGluAla----- 56  
DB 1012 AAGATGAGAGCGGAGCTGCAGACGCAAGAGAGAGAGGCTGGCGCGCGCCACGTTT 1071  
QY 57 -----AlaSerAspAlaLeuGluAlaLeuAlaAsp 67  
DB 1072 GTGTGCGGAGCTCAAGCCCGCGGTGGGAGCGCGCATGCGGTGGAGCGCGCGCGCG 1131  
QY 68 GlnThrAspAlaLeuGlnSerGluGluAlaValValLys----- 81  
DB 1132 GAGCTGAGCGGCG---CAGCGGAGAGAGCGCGCCACCGCATGGAGCGGAGCGCAAGCG 1188  
QY 82 -----AlaAspAsnAlaLaserAspAlaLeuGluAlaLeuAlaAspGluThrAspAla 99  
DB 1189 GAGCGCGCGAGAGAGCGCGCGCGCGCGCGCG---GAGCGCGCGCGAGAGAGAGAGAG 1245  
QY 100 LeuGlnSerGluGluAlaGluValAlaGlnSerAspAsnAlaLaserAspAlaTrpGlu 119  
DB 1246 CTGCTGGAGAGACTGTGGCGCC-----ACGGCGCGCGCGAGCGGAGAGAG 1290  
QY 120 LysAlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProVal--- 138  
DB 1291 CCGCGCGAGAGACCGCGCGCGCGCGCTGCTGCGCGATGGCGGTGGAGCTGTGGAC 1350  
QY 139 ValLysLysGluGluAlaGlnAsnValAsnThrLeuPro 151  
DB 1351 GTGGAGGCTGAGGCTGGCCCAAGCGGTGCAGAGCTGTGCC 1389  
RESULT 21  
US-08-762-106-6  
Sequence 6, Application US/08762106  
Patent No. 5948677  
GENERAL INFORMATION:  
APPLICANT: Jarvik, Jonathan W.  
TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
TITLE OF INVENTION: TAGGING  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Harris Brotman  
STREET: 202 Coast Blvd., Suite 111  
CITY: La Jolla  
STATE: California  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/762,106  
FILING DATE: 09-DEC-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Brotman, Harris F.  
REGISTRATION NUMBER: 35,461  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 654-2428  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1581 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-762-106-6  
Alignment Scores:  
Pred. No.: 0.161 Length: 1581

Score: 90.50 Matches: 47  
Percent Similarity: 39.88% Conservative: 22  
Best Local Similarity: 27.17% Mismatches: 41  
Query Match: 11.65% Indels: 63  
DB: 2 Gaps: 8

US-09-847-539a-6 (1-159) x US-08-762-106-6 (1-1581)

QY 35 ArgAlaIleAspGluLeuLysGlnAla----- 44  
DB 925 CGCGTGTGGACAGCTGTGGAGAGCGCGCGCTGTGCAGCCAAATCGACAGACC 984  
QY 45 IleGluAspLysGluAlaThrThrAla----- 53  
DB 985 CTGGCGGACAAAGCCCGCCAGCAGCGCCAGCTGTGCAGCGCTGTGCAGCGCC 1044  
QY 54 -----IleGluAla----- 56  
DB 1045 AAGATGAGCGCGAGCTGCAGGAGGAGCTGTGCAGCGCGCGCGCCACGTTT 1104  
QY 57 -----AlaSerSerAspAlaLeuGluAlaLeuAlaAsp 67  
DB 1105 GTGCTGCGCGAGCTCAAGCCCGCGCTGTGCAGCGCTGTGCAGCGCGCGCG 1164  
QY 68 GlnThrAspAlaLeuGlnSerGluAlaAlaValValys----- 81  
DB 1165 GAGCTGACGCGC---CAGCGCGAGAGAGCGCCCAAGCCCAAGTGGAGCGCGACAGCGC 1221  
QY 82 -----AlaAspAsnAlaIleAspAspAlaLeuAlaLeuAlaAspGlnThrAspAla 99  
DB 1222 GAGCGCGCGAGAGAGCGCGCGCGCGCGC---GAGCGCGCGCGAGAGAGAGCGC 1278  
QY 100 LeuGlnSerGluGluAlaGlnValValGlnSerAspAsnAlaIleAspAlaTrpGlu 119  
DB 1279 CTGCTGAGAGGTGGCGCC-----ACGCGCGCGCGAGCGCGAGAGAG 1323  
QY 120 LysAlaIleAlaThrProIleAlaLeuAspValLysThrLysAspThrLysProVal--- 138  
DB 1324 CGCGCGAGAGAGCGCGCGCGCGCGCGCGCTGTGCAGCGAGTGGCGTGTGTGAC 1383  
QY 139 ValLysLysGluGluArgGlnAsnValAsnThrLeuPro 151  
DB 1384 GTGAGGCTGAGTGTGCCCAAGCGGTGGAGGCTGTGCC 1422

RESULT 22  
US-09-320-774-6  
; Sequence 6, Application US/09320774  
; Patent No. 6265545  
; GENERAL INFORMATION:  
; APPLICANT: Javrik, Jonathan W.  
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
; TITLE OF INVENTION: TAGGING  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harris Brotman  
; STREET: 202 Coast Blvd., Suite 111  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/320,774  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/762,106  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:

NAME: Brotman, Harris F.  
REGISTRATION NUMBER: 35,461  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 654-2428  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1581 base pairs  
type: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-320-774-6

Alignment Scores:  
Pred. No.: 0.161 Length: 1581  
Score: 90.50 Matches: 47  
Percent Similarity: 39.88% Conservative: 22  
Best Local Similarity: 27.17% Mismatches: 41  
Query Match: 11.65% Indels: 63  
DB: 4 Gaps: 8

US-09-847-539a-6 (1-159) x US-09-320-774-6 (1-1581)

QY 35 ArgAlaIleAspGluLeuLysGlnAla----- 44  
DB 925 CGCGTGTGGACAGCTGTGGAGAGCGCGCGCTGTGCAGCCAAATCGACAGACC 984  
QY 45 IleGluAspLysGluAlaThrThrAla----- 53  
DB 985 CTGGCGGACAAAGCCCGCCAGCAGCGCCAGCTGTGCAGCGCTGTGCAGCGCC 1044  
QY 54 -----IleGluAla----- 56  
DB 1045 AAGATGAGCGCGAGCTGCAGGAGGAGCTGTGCAGCGCGCGCGCCACGTTT 1104  
QY 57 -----AlaSerSerAspAlaLeuGluAlaLeuAlaAsp 67  
DB 1105 GTGCTGCGCGAGCTCAAGCCCGCGCTGTGCAGCGCTGTGCAGCGCGCGCGC 1164  
QY 82 -----AlaAspAsnAlaIleAspAspAlaLeuAlaLeuAlaAspGlnThrAspAla 99  
DB 1222 GAGCGCGCGAGAGAGCGCGCGCGCGCGC---GAGCGCGCGCGAGAGAGAGAGCGC 1278  
QY 100 LeuGlnSerGluGluAlaGlnValValGlnSerAspAsnAlaIleAspAlaTrpGlu 119  
DB 1279 CTGCTGAGAGGTGGCGCC-----ACGCGCGCGCGAGCGCGAGAGAG 1323  
QY 120 LysAlaIleAlaThrProIleAlaLeuAspValLysThrLysAspThrLysProVal--- 138  
DB 1324 CGCGCGAGAGAGCGCGCGCGCGCGCGCGCTGTGCAGCGAGTGGCGTGTGTGAC 1383  
QY 139 ValLysLysGluGluArgGlnAsnValAsnThrLeuPro 151  
DB 1384 GTGAGGCTGAGTGTGCCCAAGCGGTGGAGGCTGTGCC 1422

RESULT 23  
US-08-106-981-1  
; Sequence 1, Application US/08106981  
; Patent No. 573149  
; GENERAL INFORMATION:  
; APPLICANT: SARHAN, Fathey  
; APPLICANT: SARHAN, Mario  
; APPLICANT: LALIBERT, Jean-Fran ois  
; TITLE OF INVENTION: DNA MOLECULES ENCODING FREEZING  
; TITLE OF INVENTION: TOLERANCE PROTEINS IN GRAMINAE  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:







```

1      SOFTWARE: Patentin Release #1.0, Version #1.25
2
3      CURRENT APPLICATION DATA:
4      APPLICATION NUMBER: US/08/214,164
5      FILING DATE: 17-MAR-1994
6      CLASSIFICATION: 424
7
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: US 07/656,773
10     FILING DATE: 15-FEB-1991
11     CLASSIFICATION: 424
12
13     ATTORNEY/AGENT INFORMATION:
14     NAME: Beistresser, Jerry W.
15     REGISTRATION NUMBER: 22,651
16     REFERENCE/DOCKET NUMBER: 6102-137
17
18     TELECOMMUNICATION INFORMATION:
19     TELEPHONE: (703) 415-0810
20     TELEFAX: (703) 521-0813
21
22     NAME: LUKPAT WASHINGTON
23
24     INFORMATION FOR SEQ ID NO: 1:
25     SEQUENCE CHARACTERISTICS:
26     LENGTH: 2085 base pairs
27     TYPE: nucleic acid
28     STRANDEDNESS: double
29     TOPOLOGY: linear
30
31     MOLECULE TYPE: DNA (genomic)
32
33     FEATURE:
34     NAME/KEY: Intron
35     LOCATION: 1..1983
36
37     FEATURE:
38     NAME/KEY: CDS
39     LOCATION: 127..1983
40
41     US-08-214-164-1

```



Oy	115	-----SerAspAlaIrrpGluLysAlaAlaThrProIleLeuLeuAspValLysThr	132
		:::	:::
Db	994	TACTTTAAGAAAGCTTTAGAGAAACATTCCTGCTCAAAAAAGCTGATTAACAAAAACT	105
Oy	133	Lys---AspThrLysProValValLysLysGluArg	144
		:::	:::
Db	1054	GAACCTGACCTTAAGAAAGCAGTTAATGACCCAGAAAAA	1092
RESULT 30			
US-08-468-718-1			
Sequence 1, Application US/08468718			
Patent No. 5871943			
GENERAL INFORMATION:			
APPLICANT: Briles, David E			
APPLICANT: Yother, Janet L			
APPLICANT: McDaniel, Larry S			
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL			
NUMBER OF SEQUENCES: 6			
TITLE OF INVENTION: PROTEIN			
NUMBER OF SEQUENCES: 6			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Shoemaker and Mattare, Ltd			
STREET: Suite 1203, 2001 Jefferson Davis Highway			
CITY: Arlington			
STATE: Virginia			
COUNTRY: U.S.A.			
ZIP: 22202-0286			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.25			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/468,718			
FILING DATE:			
CLASSIFICATION: 435			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US/08/072,068			
FILING DATE: 03 JUNE 1993			
APPLICATION NUMBER: US/07/835,698			
FILING DATE: 12-FEB-1992			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US/07/656,773			
FILING DATE: 15-FEB-1991			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (703) 415-0810			
TELEFAX: (703) 521-0378			
TELEX: LUKPAT WASHINGTON			
INFORMATION FOR SEQ ID NO: 1:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 2085 base pairs			
TYPE: nucleic acid			
STRANDEDNESS: double			
TOPOLOGY: linear			
MOLECULE TYPE: protein			
HYPOTHETICAL: NO			
ANTI-SENSE: NO			
ORIGINAL SOURCE:			
ORGANISM: Streptococcus pneumoniae			
STRAIN: rx1			
IMMEDIATE SOURCE:			
CLONE: JY2008			
FEATURE:			
NAME/KEY: Intron			
LOCATION: 1..2085			
FEATURE:			
NAME/KEY: CDS			
LOCATION: Join(127..1983, 1987..1992, 1996..2007, 2011			
LOCATION: ..2025, 2029..2031, 2035..2085)			
US-08-468-718-1			
Alignment Scores: 0.263 Length: 2085			
red. No.:			

Score: 90.00 Matches: 37  
 Percent Similarity: 39.22% Conservative: 23  
 Best Local Similarity: 24.18% Mismatches: 61  
 Query Match: 11.58% Indels: 32  
 DB: 2 Gaps: 4

US-09-847-539a-6 (1-159) x US-08-468-718-1 (1-2085)

```

QY 24 AlaProgluylsleuAlaLeuArgAsnGluArgAla---lleaspgluLeuLysLys 42
DB 634 GCACGAGAACTTACTTAAATACTAGAGAGAGCTAAAGCAAAATTAGAGAGGCTGAGAAA 693
QY 43 GluAlaIleGluAspLysGluAlaThrThraIleGluAlaIleSerSerAspAlaLeu 62
DB 694 AAAGCTCTGAGAGCAAAACCAAAAGTGATGCTGAGAGAGCTGCTCAAGCTAAATC 753
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal----- 79
DB 754 GCTGAATTGGAATCATGATTGACTAGACACAGAGCTCAAGAGATTGATGAGTCT 813
QY 80 -----ValLysAla 82
DB 814 GAATCAGAGATTATGCTAAAGAGGTTCCGCTCTTCAATCTAAATTGATGCC 873
QY 83 AspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102
DB 874 AAAAAGCTTAACTATCAAAACTTGAAGAGTTAGTGATTAAGATTGAGATTGAGCGCT 933
QY 103 GluGluAlaGluValValGlnSerAspAlaAla----- 114
DB 934 GAATTTGCAAAAGCTTGAAGATCAACTTAAAGCTGCTGAGAAAACATTAATGTGAAGAC 993
QY 115 -----SerAspAlaTrpGluLysAlaIleThrProIleAlaLeuAspValLysLysThr 132
DB 994 TACTTTAAAGAGCTTGAAGAAACCTATTGCTGCTAAAGAGCTGAATTAAGAAAAAAGCT 1053
QY 133 Lys---AspThrLysProValValLysGluLysLysLys 144
DB 1054 GAAGCTGACCTTAAGAAAGCAAGTTAATGAGCCAGAAAAA 1092

```

RESULT 31  
 US-08-246-636-1  
 : Sequence 1, Application US/08246636  
 : Patent No. 5965141  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: Biles, David E  
 : APPLICANT: Yocher, Janet L  
 : APPLICANT: McDaniel, Larry S  
 : APPLICANT: Wu, Hong-Yin  
 : TITLE OF INVENTION: EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE  
 : TITLE OF INVENTION: PROTEIN A  
 : NUMBER OF SEQUENCES: 5  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Shoemaker and Matlare, Ltd  
 : STREET: Suite 1203, 2001 Jefferson Davis Highway  
 : City: Arlington  
 : STATE: Virginia  
 : COUNTRY: U.S.A.  
 : ZIP: 22202-0286  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : OPERATING SYSTEM: IBM PC compatible  
 : SOFTWARE: PatentIn Release #1.0, Version #1.25  
 :  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/246,636  
 : FILING DATE: 20-MAY-1994  
 : CLASSIFICATION: 435  
 :  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 07/656,773  
 : FILING DATE: 15-FEB-1991  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 07/835,698

FILING DATE: 12-FEB-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/048,896  
 FILING DATE: 20-APR-1993  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0813  
 TELEX: LUKPAT WASHINGTON  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2085 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Streptococcus pneumoniae  
 STRAIN: Rxi  
 IMMEDIATE SOURCE:  
 CLONE: JY4313  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 1..2085  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: join(127..1984)  
 US-08-246-636-1

Alignment Scores:  
 Score: 0.263 Length: 2085  
 Percent Similarity: 39.22% Matches: 37  
 Best Local Similarity: 24.18% Conservative: 23  
 Query Match: 11.58% Mismatches: 61  
 DB: 2 Indels: 32  
 Gaps: 4

US-09-847-539a-6 (1-159) x US-08-246-636-1 (1-2085)

```

QY 24 AlaProgluylsleuAlaLeuArgAsnGluArgAla---lleaspgluLeuLysLys 42
DB 634 GCACGAGAACTTACTTAAATACTAGAGAGAGCTAAAGCAAAATTAGAGAGGCTGAGAAA 693
QY 43 GluAlaIleGluAspLysGluAlaThrThraIleGluAlaIleSerSerAspAlaLeu 62
DB 694 AAAGCTCTGAGAGCAAAACCAAAAGTGATGCTGAGAGAGCTGCTCAAGCTAAATC 753
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal----- 79
DB 754 GCTGAATTGGAATCATGATTGACTAGACACAGAGCTCAAGAGATTGATGAGTCT 813
QY 80 -----ValLysAla 82
DB 814 GAATCAGAGATTATGCTAAAGAGGTTCCGCTCTTCAATCTAAATTGATGCC 873
QY 83 AspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102
DB 874 AAAAAGCTTAACTATCAAAACTTGAAGAGTTAGTGATTAAGATTGAGATTGAGCGCT 933
QY 103 GluGluAlaGluValValGlnSerAspAlaAla----- 114
DB 934 GAATTTGCAAAAGCTTGAAGATCAACTTAAAGCTGCTGAGAAAACATTAATGTGAAGAC 993
QY 115 -----SerAspAlaTrpGluLysAlaIleThrProIleAlaLeuAspValLysLysThr 132
DB 994 TACTTTAAAGAGCTTGAAGAAACCTATTGCTGCTAAAGAGCTGAATTAAGAAAAAAGCT 1053
QY 133 Lys---AspThrLysProValValLysGluLysLysLys 144
DB 1054 GAAGCTGACCTTAAGAAAGCAAGTTAATGAGCCAGAAAAA 1092

```

RESULT 32



```

: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus pneumoniae
: STRAIN: Rxi
: IMMEDIATE SOURCE:
: CLONE: JY4313
: FEATURE:
: NAME/KEY: Intron
: LOCATION: 1..2085
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(127..1984)
: US-08-319-795-1

Alignment Scores:
Pred. No.: 0.263 Length: 2085
Score: 90.00 Matches: 37
Percent Similarity: 39.22% Conservative: 23
Best Local Similarity: 24.18% Mismatches: 61
Query Match: 11.58% Indels: 32
DB: 2 Gaps: 4

US-09-847-539a-6 (1-159) x US-08-319-795-1 (1-2085)
Oy 24 Alaprogilulsleualaleuargansglunargala---ileaspgluleuyls 42
Db 634 GCACCAACTTACTTAAAAAACTGAGAGAGCTTAAGCAAAATTGAGAGGCTGAGAAA 693
Oy .43 glualalegluasplysgluatlrthralaleglualaleuaseraspsalaleu 62
Db 694 AAAGCTACTGAGCCAAACAAAGGTGATGCTGAGAGAGTCCCTCAAGCTAAATC 753
Oy 63 glualaleuaspglnthraspsalaleuaglinsergluualaleuaval----- 79
Db 754 GCTGAATTGGAAATCAAGTTCATGACTGACACAAAGCTCAAGAGATTGATGACTCT 813
Oy 80 -----vallysala 82
Db 814 GAATCAGACATTTGCTAAAGAGTTCCGTCCTCTTCATCTTAATTGGATGCC 873
Oy 83 Aspsanlaleaseraspsalaleuaglualaleuaspglnthraspsalaleu 102
Db 874 AAAAAGCTAAACTATCAAACTGAGAGCTTAAGTATGATGATGATGATGATGATGAT 933
Oy 103 glunlaleglualaleuaglinseraspsanlaleu----- 114
Db 934 GAAATTCGAAACCTGGAAGATCAACTTAAGCTGCTGGAAGAAACAATTAATGTAAGAC 993
Oy 115 -----seraspsalaleuaspglnthraspsalaleuaspvallysylsthr 132
Db 994 TACTTTAAAGAGGTTTAAAGAAACTATTTGCTCTTAAAAAAAGCTGAATTTGAAAAA 1053
Oy 133 Lys---AspThrLysProvalValLysLysglunarg 144
Db 1054 GAAGCTGACCTTAAGAAAAGCAGTTAATGAGCCAGAAAAA 1092

RESULT 34
US-08-468-985-1
: Sequence 1, Application US/08468985
: Patent No. 5997882
: GENERAL INFORMATION:
: APPLICANT: Biller, David E.
: APPLICANT: Yohner, Janet L.
: APPLICANT: McDaniel, Larry S.
: TITLE OF INVENTION: Epitopic Regions of Pneumococcal Surface
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheomaker and Mattare, Ltd.
: STREET: 1203 Crystal Plaza Bldg. 1, 2001 Jefferson
: CITY: Arlington
: STATE: Virginia
```

```

: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/468,985
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/319,795
: FILING DATE:
: APPLICATION NUMBER: US 08/246,636
: FILING DATE: 20-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/048,896
: FILING DATE: 20-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/835,698
: FILING DATE: 12-FEB-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/656,773
: FILING DATE: 15-FEB-1991
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2085 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus pneumoniae
: STRAIN: Rxi
: IMMEDIATE SOURCE:
: CLONE: JY4313
: FEATURE:
: NAME/KEY: Intron
: LOCATION: 1..2085
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(127..1984)
: US-08-468-985-1

Alignment Scores:
Pred. No.: 0.263 Length: 2085
Score: 90.00 Matches: 37
Percent Similarity: 39.22% Conservative: 23
Best Local Similarity: 24.18% Mismatches: 61
Query Match: 11.58% Indels: 32
DB: 2 Gaps: 4

US-09-847-539a-6 (1-159) x US-08-468-985-1 (1-2085)
Oy 24 Alaprogilulsleualaleuargansglunargala---ileaspgluleuyls 42
Db 634 GCACCAACTTACTTAAAAAACTGAGAGAGCTTAAGCAAAATTGAGAGGCTGAGAAA 693
Oy .43 glualalegluasplysgluatlrthralaleglualaleuaseraspsalaleu 62
Db 694 AAAGCTACTGAGCCAAACAAAGGTGATGCTGAGAGAGTCCCTCAAGCTAAATC 753
Oy 63 glualaleuaspglnthraspsalaleuaglinsergluualaleuaval----- 79
Db 754 GCTGAATTGGAAATCAAGTTCATGACTGACACAAAGCTCAAGAGATTGATGACTCT 813
Oy 80 -----vallysala 82
```





QY 43 GlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleAspSerAspAlaLeu 62  
Db 694 AAAGCTACTAGCAACCAAAAGGATGCTGAAGAGCTGCTCTCAAGCTAAATC 753  
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal----- 79  
Db 754 GGTGAATTCGAATCATCAAGTTCACTAGACTGACACAGCTCAAGAGATTGATGACT 813  
QY 80 -----ValLysAla 82  
Db 814 GAATCAAGATTATGCTAAAGAGCTTCCGCTCTCTCAATCAATTAATGATGCC 873  
QY 83 AspAsnAlaIleAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102  
Db 874 AAAAAGCTTAAGCTTAAGAGAGCTTGAAGAGTTAGATGATGATGATGAGCT 933  
QY 103 GluGluAlaGluAlaValGlnSerAspAsnAlaIle----- 114  
Db 934 GAATTCGAAGCTTGAAGCTTAAGCTTGAAGAGCTTGAAGAGCTTGAAGAGCT 993  
QY 115 -----SerAspAlaIleGluAlaIleThrProIleAlaLeuAspValLysThr 132  
Db 994 TACTTTAAAGAGCTTGAAGAGCTTGAAGAGCTTGAAGAGCTTGAAGAGCT 1053  
QY 133 Lys-----AspThrLysProValValLysLysGluGluArg 144  
Db 1054 GAAGCTGACCTTGAAGAGCTTGAAGAGCTTGAAGAGCTTGAAGAGCT 1092

RESULT 37  
US-08-937-271-9  
Sequence 9, Application US/08937271  
Patent No. 6063386  
GENERAL INFORMATION:  
APPLICANT: Dale, James B.  
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN  
TITLE OF INVENTION: VACCINE  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/937,271  
FILING DATE: 15-SEP-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Roseman, Stephen J.  
REGISTRATION NUMBER: 43,058  
REFERENCE/DOCKET NUMBER: 481112, 405C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 918 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pyogenes

FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..918  
US-08-937-271-9

Alignment Scores:  
Pred. No.: 0.134 Length: 918  
Score: 88.50 Matches: 44  
Percent Similarity: 37.02% Conservative: 23  
Best Local Similarity: 24.31% Mismatches: 69  
Query Match: 11.39% Indels: 45  
DB: 3 Gaps: 5

US-09-847-539a-6 (1-159) x US-08-937-271-9 (1-918)

QY 19 AsnLeuGluGlnAlaProGluLysLeuAlaLeuArgAsnGluArgAlaIleAsp 38  
Db 313 AACAAATTTCAGAAAGAGTGGCAAA---GGCTTCGGCGTATTTAGACGCATCAGCT 369  
QY 39 GluLeuLysLysGluAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleSer 58  
Db 370 GAAGCTGAAGAGCAATTAGAGCTGAACACAAACTTGAAGAGCAAAAGATTCA 429  
QY 59 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 78  
Db 430 GAAGCAAGCTGCAAAAGGCTTCCGCTGATTTAGACCATCTACCTGACGTAAGAAACA 489  
QY 79 ValValLysAlaAspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 98  
Db 490 GTTGAAAGAGCTTGAAGAGAGCAAAACGCAAAATTACCTGCTTGAAGAACTTAACAA 549  
QY 99 AlaLeu-----GlnSerGluGluAlaGluValValGlnSerAsp 111  
Db 550 GAGCTTGAAGAGCAAAAGTAATTAACAGAAAGAAAGAGCTGAGCTCAAGCAAACTT 609  
QY 112 AsnAlaIleSerAspAlaIleGluLysAlaIleThrProIleAlaLeuAspValLysThr 131  
Db 610 GAAGCAGAGCAAAAGAGCTGCAAAAGCAATTAAGCAAAAGAGCTGCAAAAGCTTGA 669  
QY 132 -----ThrLysAspThrLysPro----- 137  
Db 670 CTAGAGCTGGAAGAGCATCACTCAACAAACCTGATACAAACGAGAAAGAGCT 729  
QY 138 -----Val 138  
Db 730 GTTCAGCTAAAGTCAAGCACCAAGCAGTACAAACCAAAACCAAGCAGCACCA 789  
QY 139 ValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGluGluGlnSerAsn 158  
Db 790 ATGAAGGAAGCAATTAAGAGCAG-----TTACCATCAACAGGTGAAGAGCCTAAC 837  
QY 159 Pro 159  
Db 838 CCA 840

RESULT 38  
US-07-813-584A-1  
Sequence 1, Application US/07813584A  
Patent No. 5352588  
GENERAL INFORMATION:  
APPLICANT: Fischetti, Vincent A.  
APPLICANT: Bessen, Debra E.  
TITLE OF INVENTION: No. 5352588e1 Immunoglobulin A Binding Protein  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kiltie Murray  
STREET: 98 Cutler Mill Road  
CITY: Great Neck  
STATE: NY  
COUNTRY: USA  
ZIP: 11021  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

```

: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/813,584A
: FILING DATE: 19911224
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Murray, Kittle
: REGISTRATION NUMBER: 30,246
: REFERENCE/DOCKET NUMBER: RU-100
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 516-482-1990
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1560 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2..40
: NAME/KEY: CDS
: LOCATION: 252..1367
: US-07-813-584A-1

Alignment Scores:
Pred. No.: 0.268 Length: 1560
Score: 88.50 Matches: 37
Percent Similarity: 40.56% Conservative: 21
Best Local Similarity: 25.87% Mismatches: 46
Query Match: 11.39% Indels: 39
DB: 1 Gaps: 5

US-09-847-539a-6 (1-159) x US-07-813-584A-1 (1-1560)
QY 17 LeuThAsnLeuLyuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGuaGala 36
DB 984 GTTACATCTGAGTTAAACACAGCAAAAGCTCAACTCTCAGCGCTGACAGAAAGTAGAAA 1043
QY 37 TLeaSPGluLeuLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 56
DB 1044 TTATCGAAAAAGAAAAAGCTGAGTTACAGCAAAA-----CTAGATGCA 1088
QY 57 AlaSerSerAspAlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 76
DB 1089 CAAGGAAAGCCCTCAAGACACATTAGCAAAACAAACT-----GACGAG 1133
QY 77 AlaAlaValAlaLysAlaAspAsnAlaAlaSerAspAlaLeuGlnAlaLeuAlaAspGln 96
DB 1134 CTTGCAAAAAGTAAGAGCTGAAAAAGCGGCGCTTCAAAAAACACCTGCT-----1181
QY 97 ThrAspAlaLeuGlnSerGluGlnAlaGluValAlaGlnSerAspAsnAlaAlaSerAsp 116
DB 1182 -----ACCAACCGCTAATTAAGAAAGCTCAGTACAGCTCTCTCAAGCA 1226
QY 117 AlaTPGluLysAlaAlaThrProIleAlaLeuAspValLysThrLysAspThrLys 136
DB 1227 GCT-----ACMAA 1235
QY 137 ProValValLysGluGlnUargGlnAsnValAsnThrLeuProThrThrGluGlnGlu 156
DB 1236 CTTACCCAAATAAAGAAATGAGA-----TCACATTTACCGTCACAGCGCAAGCA 1286
QY 157 SerAsnPro 159
DB 1287 CCTAACCCA 1295

RESULT 39
US-08-330-515-1
: Sequence 1, Application US/08330515
```

```

: Patent No. 5556944
: GENERAL INFORMATION:
: APPLICANT: Fischetti, Vincent A.
: APPLICANT: Bessen, Debra E.
: TITLE OF INVENTION: No. 5556944e1 Immunoglobulin A Binding Protein
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Saliwanchik & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/330,515
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/813,584
: FILING DATE: 24-DEC-1991
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: RU-100.D1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1560 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2..40
: NAME/KEY: CDS
: LOCATION: 252..1367
: US-08-330-515-1

Alignment Scores:
Pred. No.: 0.268 Length: 1560
Score: 88.50 Matches: 37
Percent Similarity: 40.56% Conservative: 21
Best Local Similarity: 25.87% Mismatches: 46
Query Match: 11.39% Indels: 39
DB: 1 Gaps: 5

US-09-847-539a-6 (1-159) x US-08-330-515-1 (1-1560)
QY 17 LeuThAsnLeuLyuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGuaGala 36
DB 984 GTTACATCTGAGTTAAACACAGCAAAAGCTCAACTCTCAGCGCTTACAGAAAGTAGAAA 1043
QY 37 TLeaSPGluLeuLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 56
DB 1044 TTATCGAAAAAGAAAAAGCTGAGTTACAGCAAAA-----CTAGATGCA 1088
QY 57 AlaSerSerAspAlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 76
DB 1089 CAAGGAAAGCCCTCAAGACACATTAGCAAAACAAACT-----GACGAG 1133
QY 77 AlaAlaValAlaLysAlaAspAsnAlaAlaSerAspAlaLeuGlnAlaLeuAlaAspGln 96
DB 1134 CTTGCAAAAAGTAAGAGCTGAAAAAGCGGCGCTTCAAAAAACACCTGCT-----1181
```

```

QY 97 ThrAspAlaLeuGlnSerGluGluValIaIGlnSerAspAsnAlaIaSerAsp 116
Db 1182 -----ACCAACCAAGCTAATAAGAAAGATCAGGTAGAGCTGCTCAACA 1226
QY 117 AlaTrpGluLysAlaIaThrProIleAlaLeuAspValLysThrLysAspThrLys 136
Db 1227 GCT-----ACACAGA 1235
QY 137 ProValValLysGluGluGlnArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156
Db 1236 CCTAGCCAAATTAAGAAATGAGA-----TCACAAATACCGACAGCGGAGCA 1286
QY 157 SerAspPro 159
Db 1287 GCTAACCCA 1295

RESULT 40
US-08-533-669A-7
; Sequence 7, Application US/08533669A
; Patent No. 5834592
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,669A
; FILING DATE: 22-SEP-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ. ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1771 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1698
; US-08-533-669A-7

Alignment Scores:
Pred. No.: 0.316 Length: 1771
Score: 88.50 Matches: 47
Percent Similarity: 45.34% Conservative: 26
Best Local Similarity: 29.19% Mismatches: 70
Query Match: 11.39% Indels: 18
DB: 2 Gaps: 6

US-09-847-539A-6 (1-159) x US-08-533-669A-7 (1-1771)
QY 8 ProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlyIu 27
Db 6 CCGGTCACAGCCCTCGAGAGGAGGAGCGGTCCTCCGCGAGAGCTGGAGGCGGCGGAGGA 65

```

```

QY 27 sIeuAlaLeuArgAsnGluGluValIaIaLeuAspGluLeuLysGluAlaIleGluAs 47
Db 66 GCGGCCCCCTGGATGTCTATGCATGCGCGCGAGCCCGGTCTCCAGCCCTCGAGGA 125
QY 47 p-----LysGluAlaThrThrAlaIleGluAlaIaSerSerAspAla 62
Db 126 GCGAGCGCGTCCTCGCGCGAGCTGGAGAGGCGGAGGAGGCGCGCTGATGTCTAT 185
QY 62 uGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIaValLysAl 82
Db 186 GCATGGG---GCCAGAGAGCCCGCGTCCAGGCTCTGAGAGAGGAGCGGCTCCGCGC 242
QY 82 aAsp-----AsnAlaIaSerAspAla-----LeuGluAlaLeu-----AlaAspGln 97
Db 243 GGAAGCTGGAGGCTGCGAGAGGCGCGCGCTGAGAGCCATGCAGAGCGGAGGCG 302
QY 97 rAspAlaLeuGlnSerGluGluValIaIGlnSerAspAsnAlaIaSerAspAl 117
Db 303 CCGCTCCAGCCCTCGAGAGAGCGCGCTCTCCGCGGAG----- 345
QY 117 aTrpGluLysAlaIaThrProIleAlaLeuAspValLysThrLysAspThrLysPr 137
Db 346 -CTGAGAGAGAGCGAGAGCGCGCGCGCTGATGCATGCAGCGCGAGAGCGCGCG 404
QY 137 oValValLysGluGluGlnArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 157
Db 405 CGTCCAGCCCTCGAGAGGAGCGCGCTCTCCGCGAGCTGGAGAGGCGGAGGAGGC 464
QY 157 r 157
Db 465 G 465

```

Search completed: October 13, 2002, 06:11:42  
 Job time : 67.016 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:48:12 : Search time 224.471 seconds  
(without alignments)  
1216.148 million cell updates/sec

Title: US-09-847-539A-6  
Perfect score: 777  
Sequence: 1 VDSPIEQPRIPNGTTLTNL.....KKERQNVNLTPTTGESNP 159

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ -p2n.model -DEV-xlh  
-O/cg2\_1/USPTO.spool/US09847539/runtat\_10102002\_093105\_4982/app-query.fasta.1.526  
-DB-N.Geneseq\_032802 -QFMT-fastap -SUFFIX-std.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE-pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE-LOCAL -OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MNLBN=0 -MAXLEN=200000000  
-USER-US09847539.scgn\_1.1.48-enumat\_10102002\_093105\_4982 -NCPU=6 -ICPU=3  
-NO\_XLPY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV-TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
N\_Geneseq\_032802 : \*  
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT : \*  
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT : \*  
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT : \*  
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT : \*  
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT : \*  
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT : \*  
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT : \*  
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT : \*  
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT : \*  
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT : \*  
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT : \*  
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT : \*  
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT : \*  
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT : \*  
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT : \*  
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT : \*  
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT : \*  
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT : \*  
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT : \*  
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT : \*  
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT : \*  
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT : \*  
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT : \*  
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	777	100.0	654	21 AAD00560	Streptococcus pyog
2	777	100.0	764	21 AAD00559	Streptococcus pyog
3	760	97.8	504	21 AAD00564	Streptococcus pyog
4	750	96.5	469	21 AAD00562	Streptococcus pyog
5	694	89.3	853	21 AAD00563	Streptococcus pyog
6	692	89.1	777	21 AAD00561	Streptococcus pyog
7	188	24.2	1949	11 AAO06018	Sequence encoding
8	188	24.2	1950	10 AAN91093	Protein G gene. S
9	188	24.2	1950	10 AAN91094	Protein G gene. S
10	188	24.2	1950	11 AAO06017	Sequence encoding
11	188	24.2	1950	12 AAO10001	Sequence encoding
12	188	24.2	1950	11 AAO06464	Streptococcus Prot
13	187	24.1	2383	11 AAO06019	Sequence encoding
14	187	24.1	2383	12 AAO10002	Sequence encoding
15	187	24.1	2383	15 AAO75036	Streptococcus Prot
16	135	19.9	1376	21 AAZ47931	Streptococcus stra
17	139	17.9	798	11 AAO91099	Protein G gene var
18	139	17.9	798	11 AAO06009	Sequence encoding
19	139	17.9	798	12 AAO10007	Type 4 GX7809 prot
20	139	17.9	798	15 AAO64648	IgG-binding Strept
21	116	14.9	501	21 AAO45707	Arabidopsis thalia
22	114.5	14.7	1535	16 AAO68080	S. dysgalactiae ma
23	114.5	14.7	1535	16 AAO89197	S. dysgalactiae ma
24	112	14.4	1469	8 AAN70757	Sequence encoding
25	111	14.3	2121	23 ABL10037	Drosophila melanog
26	110	14.2	3212	20 AAX13365	Enterococcus faeca
27	109.5	14.1	812	21 AAC38790	Arabidopsis thalia
28	109	14.0	1128	21 AAC46811	Arabidopsis thalia
29	108	13.9	4086	23 ABL11003	Drosophila melanog
30	108	13.9	6782	20 AAX11002	Enterococcus faeca
31	107.5	13.8	486	20 AAX13883	Rabbit enteropatho
32	105.5	13.6	989	18 AAT91293	Synthetic protein
33	105	13.5	627	16 AAO84886	S. dysgalactiae m1
34	105	13.5	2526	16 AAO89196	Arabidopsis thalia
35	104.5	13.4	575	21 AAC52179	Staphylococcus aur
36	104	13.4	7434	23 AAS52179	Staphylococcus aur
37	104	13.4	7437	23 AAS52178	Staphylococcus aur
38	103.5	13.3	6228	23 AAS52178	Staphylococcus aur
39	103.5	13.3	6561	23 AAS55178	Staphylococcus aur
40	103.5	13.3	6744	13 AAO29471	Extracellular fact
41	103	13.3	7104	23 AAS51998	Staphylococcus aur
42	103	13.2	7107	23 AAS54654	Staphylococcus aur
43	102.5	13.2	1005	23 AAS92880	DNA encoding novel
44	102	13.1	603	16 AAO84887	Chimeric syntheti
45	102	13.1	8155	18 AAV74374	Staphylococcus aur

## ALIGNMENTS

RESULT 1  
AAD00560  
ID AAD00560 standard; DNA; 654 BP.

XX AAD00560;

CT 29-AUG-2000 (first entry)

XX Streptococcus pyogenes strain SF370 GRAB protein coding region.

XX GRAB protein; protein G related alpha2m binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

XX immune response; Streptococcus pyogenes infection; ds.

XX Streptococcus pyogenes.

XX Key Location/Qualifiers  
CDS 1..654

```

FT      /+tag- a
FT      /product= "GRAB protein"
FT      1...99
FT      sig_peptide
FT      /+tag- b
FT      mat_peptide
FT      100..651
FT      /+tag- c
FT      /product= "Mature GRAB protein"
FT
XX      WO200026240-A2.
XX
XX      11-MAY-2000.
XX
XX      02-NOV-1999; 99WO-GB03631.
XX
XX      02-NOV-1998; 98GB-0023975.
XX
XX      (ACTI-) ACTINOVA LTD.
XX
XX      Bjorck LH, Rasmussen M;
XX
XX      WPI: 2000-365572/31.
XX      P-PSDB: AAY71042.
XX
XX      New alpha2m binding protein for generating a protective immune response
XX      to group A streptococcus and purifying the binding protein
XX
XX      Claim 13; Page 63; 67pp; English.
XX
XX      The patent discloses a new family of proteins termed GRAB (protein G
XX      related alpha2m binding protein) from Streptococcus pyogenes which have
XX      the ability to bind alpha2-macroglobulin (alpha2m) and show homology to
XX      protein G of group G Streptococcus. GRAB protein and peptides derived
XX      from it are used in vaccine compositions for generating a protective
XX      immune response against group A Streptococcus. Antibodies against GRAB
XX      are useful for treating Streptococcus pyogenes infections. The protein
XX      is also useful for purifying Streptococcus pyogenes from a sample. The present sequence
XX      is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.
XX
XX      Sequence 654 BP; 234 A; 118 C; 143 G; 159 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.: 1 09e-74 Length: 654
XX      Score: 777.00 Matches: 159
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 100.00% Indels: 0
XX      DB: 21 Gaps: 0
XX
XX      US-09-847-539A-6 (1-159) x AAD00560 (1-654)
XX
XX      1 ValAspSerProIleGluInProArgIleIleProAsnGlyThrLeuThrAsnLeu 20
XX      100 GTTGACTCAACCTTACGACGCTCGAATTATTCCTCAATGGCGAACCTTAACCTTAACTT 159
XX
XX      21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluLysArgAlaIleAspGluLeu 40
XX      160 CTGGCAATCTCCAGAAAACCTGCATTACATGAAGAAAGACGACCATGTGATGAATTA 219
XX
XX      41 LysLysGluAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleAspGluLeu 60
XX      220 AAAAACAAGCACTTATGAGGTAAGAAGCTACACAGCTATTAACACAGCAAGCTTGAGAT 279
XX
XX      61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80
XX      280 GCCTTAGAAGCATTTACCGCATCAACACAGCGCTTTACATCAAGAAAGCTGCGTTGTT 339
XX
XX      81 LysAlaAspAsnAlaIleAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
XX      340 AAAGCGGATAAGCTCTAGTCAGCGCTTAGAAGCATTTGGCGGATCAACACAGCGCTTTA 399
XX
XX      101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaIleAspAlaIleAspGluLys 120
XX      400 CAATCAGAAAGAGCTGAAGTACTTCAATCAGATTAACGCTGCTAGTGACGCTGGAAAAA 459

```

```

XX      121 AlaIleThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValIleVal 140
XX      460 GCACCAACTCCAAATCGCTTTAGATGTTAAGAAAACCTTAAGTACAAAACCTGTAAGTAAA 519
XX
XX      141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGluGluSerAsnPro 159
XX      520 AAAGAAAGAAAGACAAACGTTAATACCTTCCTCAACTGCTGAAGAGTGAACCA 576
XX
XX      RESULT 2
XX      AAD00559
XX      ID AAD00559 standard; DNA: 764 BP.
XX
XX      AC AAD00559;
XX
XX      29-AUG-2000 (first entry)
XX
XX      Streptococcus pyogenes strain SF370 GRAB protein encoding DNA.
XX
XX      GRAB protein: protein G related alpha2m binding protein; vaccine;
XX      alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX      immune response; Streptococcus pyogenes infection; ds.
XX
XX      Streptococcus pyogenes.
XX
XX      Key Location/Qualifiers
XX      CDS 60..713
XX
XX      /+tag- a
XX      /product= "GRAB protein"
XX      60..158
XX      /+tag- b
XX      159..710
XX      /+tag- c
XX      /product= "Mature GRAB protein"
XX      complement (101..124)
XX      /+tag- d
XX      /bound_molecy= "Primer 1 represented in AAD00565"
XX      complement (101..127)
XX      /+tag- e
XX      /bound_molecy= "Primer 2 represented in AAD00566"
XX      complement (160..184)
XX      /+tag- f
XX      /bound_molecy= "Primer 3 represented in AAD00567"
XX      complement (563..594)
XX      /+tag- g
XX      /bound_molecy= "Primer 4 represented in AAD00568"
XX      complement (605..626)
XX      /+tag- h
XX      /bound_molecy= "Primer 5 represented in AAD00569"
XX
XX      WO200026240-A2.
XX
XX      11-MAY-2000.
XX
XX      02-NOV-1999; 99WO-GB03631.
XX
XX      02-NOV-1998; 98GB-0023975.
XX
XX      (ACTI-) ACTINOVA LTD.
XX
XX      Bjorck LH, Rasmussen M;
XX
XX      WPI: 2000-365572/31.
XX      P-PSDB: AAY71042.
XX
XX      New alpha2m binding protein for generating a protective immune response
XX      to group A streptococcus and purifying the binding protein
XX
XX      Example 1; Fig 2B; 67pp; English.
XX
XX      The patent discloses a new family of proteins termed GRAB (protein G
XX      related alpha2m binding protein) from Streptococcus pyogenes which have
XX      the ability to bind alpha2-macroglobulin (alpha2m) and show homology to

```

CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2m from a sample. The present sequence  
 CC is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.  
 XX

Sequence 764 BP; 279 A; 131 C; 159 G; 195 T; 0 other;

# Alignment Scores:

Pred. No.:	1-31e-74	Length:	764
Score:	777.00	Matches:	159
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-847-539a-6 (1-159) x AAD00559 (1-764)

```

QY 1 ValAspSerProIleGluInProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
DB 159 GTTGACTACCTATCGAAGCCCTCGAATTAATTCCTCAATGCGGACCTTAACCTT 218
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
DB 219 CTGGCAATGCTCCAGAAAACCTGCGATTAAGTAAAGAAAGAGCCATTGATGATTA 278
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThraIleGluAlaIleAspSerAsp 60
DB 279 AAAAAACAAGCTATGAGTAAAGAGCTAAGCAGCTATAGAGAGCAGCAAGTTCAAGAT 338
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80
DB 339 GCCCTAAGACATTAAGCGGATCAACAGAGCTTACATCAAGAGAAAGCTGCGGTTGTT 398
QY 81 LysAlaAspAsnAlaIleAspAlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeu 100
DB 399 AAGCGGATTAAGCGCTAGAGAGCGCTTGAACCATTTGGCGGATCAACAGAGCTTTA 458
QY 101 GlnSerGluGluAlaGluValAlaGlnSerAspAsnAlaIleAspAlaIleAspGluLys 120
DB 459 CAATCAGAAAGAGCTGAAGTGAATCAATCAAGCTGCTAGTGAAGCGCTGCGGAAAAA 518
QY 121 AlaIleThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLys 140
DB 519 GCACCAACTCAATCGCTTAGATGTTAAGAAACTAAAGATACAAAACCTGTAGTTAAA 578
QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAspPro 159
DB 579 AAAGAAAGAAAGAGCAAAAGCTTAATACCTTCTCAAACTGCTGAAGAGCTTAACCA 635

```

## RESULT 3

AAD00564 standard; DNA; 504 BP.

AAD00564:

29-AUG-2000 (first entry)

Streptococcus pyogenes strain KTL3 partial GRAB protein encoding DNA.

GRAB protein: protein G related alpha2m binding protein; vaccine;  
 alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
 immune response; Streptococcus pyogenes infection; ds.

Streptococcus pyogenes.

Location/Qualifiers

Key 1..504  
 CDS /tag- a  
 /product- "GRAB protein"  
 /partial

PN WO200026240-A2.

XX 11-MAY-2000.

XX 02-NOV-1999; 99WO-GB03631.

XX 02-NOV-1998; 98GB-0023975.

XX (ACTT-) ACTINOVA LTD.

XX Bjorck LH, Rasmussen M;

XX WPI: 2000-365572/31.

XX P-PSDB; AAY71046.

PT New alpha2m binding protein for generating a protective immune response  
 PT to group A streptococcus and purifying the binding protein

PS Claim 13; Page 65; 67pp; English.

CC The patent discloses a new family of proteins termed GRAB (protein G  
 CC related alpha2m binding protein) from Streptococcus pyogenes which have  
 CC the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
 CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2m from a sample. The present sequence  
 CC is a DNA encoding partial GRAB protein from S. pyogenes strain KTL3.  
 CC The protein has alpha2m binding region and is useful in vaccine  
 CC composition.

XX Sequence 504 BP; 188 A; 97 C; 108 G; 111 T; 0 other;

## Alignment Scores:

Pred. No.:	5.32e-73	Length:	504
Score:	760.00	Matches:	156
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.81%	Indels:	0
DB:	21	Gaps:	0

US-09-847-539a-6 (1-159) x AAD00564 (1-504)

```

QY 1 ValAspSerProIleGluInProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
DB 34 GTTGACTACCTATCGAAGCCCTCGAATTAATTCCTCAATGCGGACCTTAACCTT 93
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
DB 94 CTGGCAATGCTCCAGAAAACCTGCGATTAAGTAAAGAAAGAGCCATTGATGATTA 153
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThraIleGluAlaIleAspSerAsp 60
DB 154 AAAAAACAAGCTATTAAGAGATTAAGAAAGCTAAGAGCTATAGAAAGCAGCAAGTTCA 213
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80
DB 214 GCCTTAGAAGCATTAAGCGGATCAACAGAGCTTACATCAAGAAAGCTGCGGTTGTT 273
QY 81 LysAlaAspAsnAlaIleAspAlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeu 100
DB 274 AAAGCGGATTAAGCGCTAGAGAGCGCTTGAACCATTTGGCGGATCAACAGAGCTTTA 333
QY 101 GlnSerGluGluAlaGluValAlaGlnSerAspAsnAlaIleAspAlaIleAspGluLys 120
DB 334 CAATCAGAAAGAGCTGAAGTGAATCAATCAAGATTAAGAGAGCGCTGCGGAAAAA 393
QY 121 AlaIleThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLys 140
DB 394 GCACCAACTCAATCGCTTAGATGTTAAGAAACTAAAGATACAAAACCTGTAGTTAAA 453
QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156

```

Db 454 AAGAGAAAGACAAACGTTATACCTTCCTCAACTGTGAAAG 501

## RESULT 4

ID AAD00562 standard; DNA: 469 BP.

AA000562;

29-AUG-2000 (first entry)

Streptococcus pyogenes strain AP1 partial GRAB protein encoding DNA.

GRAB protein; protein G related alpha2M binding protein; vaccine;

alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

Immune response; Streptococcus pyogenes infection; ds.

Streptococcus pyogenes.

Key Location/Qualifiers

CDS 1..468

11-MAY-2000.

02-NOV-1999; 99WO-GB03631.

02-NOV-1998; 98GB-0023975.

(ACT1-) ACTINOVA LTD.

Bjorck LH, Rasmussen M;

WPI: 2000-365572/31.

P-PSDB: AAY71044.

New alpha2M binding protein for generating a protective immune response

to group A streptococcus and purifying the binding protein

Claim 13; Page 64; 67pp; English.

The patent discloses a new family of proteins termed GRAB (protein G

related alpha2M binding protein) from Streptococcus pyogenes which have

the ability to bind alpha2-macroglobulin (alpha2M) and show homology to

protein G of group G Streptococcus. GRAB protein and peptides derived

from it are used in vaccine compositions for generating a protective

immune response against group A Streptococcus. Antibodies against GRAB

are useful for treating Streptococcus pyogenes infections. The protein

is also useful for purifying Streptococcus pyogenes from a sample. The protein

is a DNA encoding partial GRAB protein from S. pyogenes strain AP1.

The protein has alpha2M binding region and is useful in vaccine

composition.

Sequence 469 BP; 180 A; 90 C; 99 G; 100 T; 0 other;

Alignment Scores:

Pred. No.: 5.79e-72 Length: 469

Score: 750.00 Matches: 154

Percent Similarity: 99.35% Conservative: 0

Best Local Similarity: 99.35% Mismatches: 1

Query Match: 96.53% Indels: 0

DB: 21 Gaps: 0

US-09-847-539a-6 (1-159) x AAD00562 (1-469)

QY 2 AbspserProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 21

DB 1 GACTCACTATGAGAACAGCGCTGAAATTAATCCAAATGCGGAACCTTAATTAATCTT 60

QY 22 GLYSAsnAlaProGluLysSLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLys 41

DB 61 GGCAATGCTCCACAGAAAACGCGCATTAATGATGAAGAAAGACCATTTGATGAATAAAA 120

QY 42 LysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleSerAspAla 61

DB 121 AACAAGCTAATGAGATAGAGAAAGCTACGACGATATACAGACACAGCTTCAGATGCC 180

QY 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleValLys 81

DB 181 TTAGAACGATTACCGCATCAACAGACGCTTTCAATCAGAAACAGCTGGGTTAAAA 240

QY 82 AlaAspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101

DB 241 GCGGATTAACGCTGCTGATGACGCTTAAGAAAGCTGGGATCAACAGACGCTTTACAA 300

QY 102 SerGluGluAlaGluValValGlnSerAspAsnAlaIleSerAspAlaIleTrpGluLysAla 121

DB 301 TCAGAAAGAGCTGAGTAGTCAATCAGATTAACGCTGATGACGCTGGGAAAAAGCA 360

QY 122 AlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLysLys 141

DB 361 GCACTCCAAATCCCTTTAGATGTTAAGAAACCTAAAGATACAAACCTGATTAATAAA 420

QY 142 GluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156

DB 421 GAGAAAGACAAACGTTAATACCTTCCTACAACTGCTGTAAGAG 465

## RESULT 5

AAD00563

ID AAD00563 standard; DNA: 853 BP.

AA000563;

29-AUG-2000 (first entry)

Streptococcus pyogenes strain AP49 partial GRAB protein encoding DNA.

GRAB protein; protein G related alpha2M binding protein; vaccine;

alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

Immune response; Streptococcus pyogenes infection; ds.

Streptococcus pyogenes.

Key Location/Qualifiers

CDS 37..852

11-MAY-2000.

02-NOV-1999; 99WO-GB03631.

02-NOV-1998; 98GB-0023975.

(ACT1-) ACTINOVA LTD.

Bjorck LH, Rasmussen M;

WPI: 2000-365572/31.

P-PSDB: AAY71045.

New alpha2M binding protein for generating a protective immune response

to group A streptococcus and purifying the binding protein

Claim 13; Page 64; 67pp; English.

The patent discloses a new family of proteins termed GRAB (protein G

related alpha2M binding protein) from Streptococcus pyogenes which have

the ability to bind alpha2-macroglobulin (alpha2M) and show homology to





```

Db 61 GTTGACTACCTATCGAACAGCCCTCGAATTATTCGAATGCCGGAACCTTAATCTT 120
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluLysArgAlaIleAspGluLeu 40
Db 121 CTTGGCAATGCTCCAGAAAACTGGCATTAAGTAATGAAGAAAGGCCATTGATGATTA 180
QY 41 LysLysGlnAlaIleGluAspLysGluAlaIleThrAlaIleGluAlaIleAspSerAsp 60
Db 181 AAAAAACAAGCTATTGAGGATTAAGAAAGTACGACAGCTATAGAACACAGCAAGTTCAGAT 240
QY 61 Ala----- 61
Db 241 GCCTTAGAAGCATTTAGCGGATCAAGCAGACGCTTTACAAATCAGAACAGCTCGATAGTT 300
QY 61 ----- 61
Db 301 CAATCAGATTAAGCTCTCTAGTACGCGCTTAGAAGCATTTGGCGGATCAACAGACCTTTA 360
QY 61 ----- 61
Db 361 CAATCAGAAAGCTCGCTGTGTTAAAGCGGATTAACGCTGCTAGTACACTTTAGAAAGCA 420
QY 61 ----- 61
Db 421 TTGGCGGATCAAAACAGACCGCTTTACATCAGAAAGAGCTGCGTGTAAAGCGGATTAAC 480
QY 62 -----LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu 76
Db 481 GCTGCTAGTACACTTTAAAGCATTTGGCGGATCAACAGACGCTTTACATCAGAAAGAA 540
QY 77 AlaAlaValAlaValAlaAspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGln 96
Db 541 GCTGCGGTTGTTAAAGCGGATTAACGCTCTAGTACACTTTAGAAAGCATTTGGCGGATCAA 600
QY 97 ThrAspAlaLeuGlnSerGluGluAlaGluValAlaGlnSerAspAsnAlaIleSerAsp 116
Db 601 ACAGACGCTTTACATCAGAAAGAGCTGAAGTACTTCAATCAGATTAACCTCTACTGTAC 660
QY 117 AlaTrpGluLysAlaAlaIleThrProIleAlaLeuAspValLysLysThrLysAspThrLys 136
Db 661 GCCTGGGGAAGAGCAGCACTCCATCGCTTAGATGTTAAGAAACATAACATACAAAA 720
QY 137 ProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGlu 155
Db 721 CCTGTAGTTAAAAAGAAAGAAAGACAAACGTTAATACCTCTCTCAACAGTGTGAA 777
RESULT 7
AA006018
ID AA006018 standard: DNA; 1949 BP.
XX
AC AA006018:
XX
DT 17-JUN-1991 (first entry)
XX
DE Sequence encoding Protein G with two active sites.
XX
KW Immunoglobulin: ds.
XX
OS Streptococcus sp. Lancefield Group G strain.
XX
FH
FT Key Location/Qualifiers
FT CDS 577..1920
FT FT /*tag= a
FT FT -35_signal 465..470
FT FT /*tag= b
FT FT -10_signal 487..492
FT FT /*tag= c
FT FT RBS 565..571
FT FT /*tag= d
XX
XX US4956296-A.
XX
XX 11-SEP-1990.

```

```

XX XX 20-JUN-1988; 88US-0209236.
PF XX
PR 20-JUN-1988; 88US-0209236.
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87MO-US00329.
PR 19-JUN-1987; 87US-0063959.
XX XX
PA (GENE-) GENEX CORP.
XX
PI Fahnstock SR;
XX
XX WPI; 1990-297491/39.
DR P-PSDB; AAR07013.
XX
XX Recombinant Protein G variants - obt'd using a cloned gene
PT encoding Protein G from Streptococcus sp., used for binding
PT Immunoglobulin.
XX
XX PS Disclosure; Fig 8a-c; 48pp; English.
XX
XX CC Fragments and variants of the sequence are claimed esp. where
XX CC incorporated into a non-pathogenic host eg. E.coli, and expressed
XX CC at high levels.
XX CC The variants have a higher binding efficiency and capacity for
XX CC Immunoglobulin, and may be used for purifying, detecting and
XX CC isolating antibodies.
SQ Sequence 1949 BP; 705 A; 323 C; 398 G; 523 T; 0 other;

Alignment Scores:
Pred. No.: 8,12e-11 Length: 1949
Score: 188.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 36.85% Mismatches: 60
Query Match: 24.20% Indels: 16
Gaps: 6

US-09-847-539A-6 (1-159) x AA006018 (1-1949)
QY 1 ValAspSerProIleGluGlnProAlaIleIleProAsnGlyLysThrLeuThrAsnLeu 20
Db 676 GTTGATTCACCAATCGAATATATTCGTAAGAGTGCTGAATTAATCTAATCTT 735
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Db 736 CTGGGGAATTCAGACAGCAACACTGCTTGGCTGATGAAGAGAGTCTACAGCTGATTG 795
QY 41 LysLysGlnAlaIleGluAspLysGluAlaIleThrAlaIleGluAlaIleAspSerAsp 60
Db 796 ACAGCAGCAGCGGTACCGCATCTGCGCAGCAGCGCGAGGAAATTCCTGGCGCAGCA 855
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80
Db 856 GCTTGGGAAGCAGCCGCA--GCACGAGATGCTCTAGCAAAAGCCAAACAGATGCCCTT 912
QY 81 LysAlaAspAsn-----AlaIleSerAspAlaLeuGluAlaLeu-----AlaAsp 95
Db 913 AAAGAATTCAACAAATTTGAGTAGTACTGATTTACAAAGATCTAATCAACATGCCAA 972
QY 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValAlaGlnSerAspAsnAlaIle--- 114
Db 973 ACTGTGAAGCATTAAGAACCTTCAAGCAGCAAGTGTGATTAACAGCAAGAAAGCGCGT 1032
QY 115 ---SerAspAlaIleGlu-----LysAlaAlaIleThrProIleAlaLeu 127
Db 1033 ATTTCAGAAAGCAACAGATGCTTATCTGATTTCTTGAAATTCGCAAAACACCTGCTGAAGAT 1092
QY 128 AspValLysLysThrLys-----AspThrLysProValValLysGlu 142
Db 1093 ACTGTTAAATCAATTGAATTAGCTCAAGCTAAAGCTTACGTAACAGAGAA 1143

```

```

RESULT 8
ID AAN91093 standard; DNA: 1950 BP.
AC AAN91093;
XX
XX 04-JUL-1990 (first entry)
XX
XX Protein G gene.
XX
XX Protein G; immunoglobulin; Fc receptor; ds.
XX
XX Streptococcus sp.
XX
XX Key Location/Qualifiers
XX FT -35_signal 465..470 /*tag- c
XX FT -10_signal 487..492 /*tag- b
XX FT RBS 565..571 /*tag- d
XX FT CDS 578..1918 /*tag- a
XX
XX MO8810306-A.
XX
XX 29-DEC-1988.
XX
XX 20-JUN-1988; 88WO-0502084.
XX
XX 19-JUN-1987; 87US-0063959.
XX
XX (GENE-) GENEX CORP.
XX
XX Fahnestock SR;
XX
XX WPI: 1989-023848/03.
XX
XX P-PSDB: AAP95030.
XX
XX Cloned protein G variant genes -
XX PT expressing proteins having immunoglobulin-binding properties of
XX PT protein G and derived from Streptococcus sp.
XX
XX PS Disclosure; ; 11pp; English.
XX
XX CC Gene encodes protein G of non-pathogenic streptococcus sp. allowing
XX CC isolation of the protein and variants, useful as bacterial Fc receptors
XX CC eg in purification and detection of Abs., screening of hybridoma clones
XX CC and treatment of disease.
XX
XX SQ Sequence 1950 BP; 705 A; 323 C; 398 G; 524 T; 0 other;

Alignment Scores:
Pred. No.: 8.12e-11 Length: 1950
Score: 188.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 24.20% Indels: 16
DB: Gaps: 6

US-09-847-539a-6 (1-159) x AAN91093 (1-1950)
OY 1 ValasparProIleGluInProArgTLeIleProAsnGlyIYThrIeuThrAsnIeu 20
DB 677 GTTGATTCACCAATCGAAGATACCCCAATATTCGTAAATGCTGATTAACCTAATCTT 736
OY 21 LeuLysAlaIaIaProGluLysLeuAlaIeuArgAsnGluGuaIaGAlaIleAspGluIeu 40
DB 737 CTGGGGAATTCAGACAAACACTGCTTGGTATGAGAGAGTGTACAGCTGATTG 796
OY 41 LysLysGlnAlaIaIeGluAspLysGluAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 60
DB 797 ACACGACGACGCGTACGCCGATCTGTGACACGACGCGTAAATGCTGGGGACACA 856

```

```

OY 61 AlaleuGluAlaIeuAlaAspGlnThrAspAlaIeuInserGluGluAlaIaIaIaIaIaIaIa 80
DB 857 GCTTGGGAGACGCGCA--GCAGCAGATGCTTACGCAAAAGCCAAAGATGCCCTT 913
OY 81 LysAlaAspAsn-----AlaAlaSerAspAlaIeuGluAlaIeu-----AlaAsp 95
DB 914 AAGCAATTCACCAATATGAGTAAAGTACTATTACAGAACTCAACCAATGCCAA 973
OY 96 GlnThrAspAlaIeuInserGluGluAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 114
DB 974 ACTGTGACGCAATAAAGACCTTCAAGCAAGTGTGATCAGACGGAAGAAAGCGCT 1033
OY 115 ---SerAspAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 127
DB 1034 ATTTCAGACGCAACAGATGCTTATCTGATTTCTTGAATCGCAACACCTGCTGAGAT 1093
OY 128 AspValLysLysThrLys-----AspThrLysProValLysLysGlu 142
DB 1094 ACTGTTAAATCAATTGAAATTAGCTGAAGCTAAAGTCTTACCTAACAGAGA 1144

RESULT 9
ID AAN91094 standard; DNA: 1950 BP.
AC AAN91094;
XX
XX 04-JUL-1990 (first entry)
XX
XX Protein G gene.
XX
XX Protein G; immunoglobulin; Fc receptor; ds.
XX
XX Streptococcus sp.
XX
XX Key Location/Qualifiers
XX FT -35_signal 465..470 /*tag- c
XX FT -10_signal 487..492 /*tag- b
XX FT RBS 565..571 /*tag- d
XX FT CDS 578..1918 /*tag- a
XX
XX MO8810306-A.
XX
XX 29-DEC-1988.
XX
XX 20-JUN-1988; 88WO-0502084.
XX
XX 19-JUN-1987; 87US-0063959.
XX
XX (GENE-) GENEX CORP.
XX
XX Fahnestock SR;
XX
XX WPI: 1989-023848/03.
XX
XX P-PSDB: AAP95030.
XX
XX Cloned protein G variant genes -
XX PT expressing proteins having immunoglobulin-binding properties of
XX PT protein G and derived from Streptococcus sp.
XX
XX PS Disclosure; ; 11pp; English.
XX
XX CC Gene encodes protein G of non-pathogenic streptococcus sp. allowing
XX CC isolation of the protein and variants, useful as bacterial Fc receptors
XX CC eg in purification and detection of Abs., screening of hybridoma clones
XX CC and treatment of disease.
XX
XX SQ Sequence 1950 BP; 706 A; 324 C; 399 G; 521 T; 0 other;

```



```

KM Immunoglobulins; Ig; ds.
XX
OS Streptococcus sp GX7809.
XX
FH Key Location/Qualifiers
FT CDS 578..2248
FT /tag= a
XX
PN US4977247-A.
XX
PD 11-DEC-1990.
XX
PF 19-MAY-1989; 89US-0354264.
XX
PR 19-MAY-1989; 89US-0354264.
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87MO-0500329.
PR 19-JUN-1987; 87US-0063959.
PR 20-JUN-1988; 88US-0209236.
XX
PA (GENE-) GENEX CORP.
XX
PI Fahnstoc SR, Lee T, Wroble MH:
XX
DR WPI: 1991-006758/01.
DR P-PSDB; AAR10004.
XX
PT Immobilised protein G variants - used for detection, isolation
PT and purification. Immunoglobulin(s) and immunoglobulin fragments
XX
PS Disclosure; Fig 8; 52pp; English.
XX
CC Protein G gene product may be modified allowing the variant to be
CC immobilised and exhibit different binding profiles. The bound
CC protein is useful in purification and detection of Igs and fragments.
XX
SO Sequence 1950 BP; 706 A; 323 C; 398 G; 523 T; 0 other;

Alignment Scores:
Pred. No.: 8.12e-11 Length: 1950
Score: 188.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 24.20% Indels: 16
Gaps: 6
DB: 12

US-09-847-539a-6 (1-159) x AAQ10001 (1-1950)
QY 1 ValAspSerProIleGluInProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
DB 677 GTTGATTCACCAATCGAGATACCCCAATATTCGTAAATGGTGAATTAACTAATCTT 736
QY 21 LeuGlyAsnAlaProGlyIuLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
DB 737 CTGGGAATTCAGAGAACACCTGGCTTGGTAAATGAAGAGATGCTACAGCTGATTG 796
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaIleAspSerAsp 60
DB 797 ACACGACGACGAGGTAGCTACTGTGCGACGACGCGAGCTGAATAATGCTGGCGACGA 856
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80
DB 857 GCTTGGAGAGCGCGCA--GCAGCAGATGCTCTAGCAAAAGCCAAAGCAGATGCCCTT 913
QY 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
DB 914 AAAGAATTCAACAATATGAGTAAAGTACATTAACAAGATCTAATCAACAATGCCAAA 973
QY 96 GlnThrAspAlaLeuGlnSerGluGluAlaIleValGlnSerAspAlaIleVal 114
DB 974 ACTGTTGAAGCATMAAAGACCTTCACAGCACACTGTTGATGATCAGGAGAAAGCGCGCT 1033

```

```

QY 115 ---SerAspAlaTrpGlu-----LysAlaIleAlaThrProIleAlaLeu 127
DB 1034 ATTTCAGACAGACACAGATGCGCTATCTGATTCTTGAAATCGCAACACCTCGTGAAGAT 1093
QY 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
DB 1094 ACTGTTAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1144

RESULT 12
AAQ64644
ID AAQ64644 standard; DNA; 1950 BP.
XX
AC AAQ64644;
XX
DT 06-JAN-1995 (first entry)
XX
DE Streptococcus Protein G gene derived from strain GX7809.
XX
KM Streptococcus Protein G; variant; IgG binding activity;
KM Immunoglobulin; Lancefield Group G; bacterial Fc receptor; ss.
XX
OS Streptococcus sp. GX7809.
XX
FH Key Location/Qualifiers
FT -35_signal 465..470
FT /tag= a
FT -10_signal 487..492
FT /tag= b
FT RBS 566..572
FT /tag= c
FT CDS 578..1924
FT /tag= d
FT /product= Protein_G
XX
PN US5312901-A.
XX
PD 17-MAY-1994.
XX
PF 14-FEB-1986; 86US-0829354.
XX
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 19-JUN-1987; 87US-0063959.
PR 20-JUN-1988; 88US-0209236.
PR 19-JUN-1990; 90US-0540169.
PR 21-APR-1992; 92US-0871539.
XX
PA (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX
PI Fahnstoc SR;
XX
DR WPI: 1994-159179/19.
DR P-PSDB; AAR53290.
XX
PT New recombinant streptococcal protein G variants - useful for
PT antibody detection and purification and for therapy
XX
PS Example 2; Fig 3; 48pp; English.
XX
CC A 1.9kb HindIII fragment containing the entire coding sequence for
CC Protein G was isolated from Streptococcus GX7809. The protein G has
CC IgG-binding activity which has been localised to the B repeating
CC structure. Streptococcal Protein G variants comprising the B domains
CC are claimed.
XX
SQ Sequence 1950 BP; 706 A; 323 C; 398 G; 523 T; 0 other;

Alignment Scores:
Pred. No.: 8.12e-11 Length: 1950
Score: 188.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 24.20% Indels: 16

```





```

US-09-847-539A-6 (1-159) x AA075036 (1-2383)
OY 1 ValAspSerProIIleGIuInProAArgIleIleProAsnGlyGlyThrIleuThrAsnLeu 20
DB 676 GTTGACTCAGCAATGCAACATACCCCAATTATTCGTAATGCTGCTCAATTAATTAATCTTT 735
OY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
DB 736 CTGGGAATTCAGAGCAACACTGCTTGGCTAATGAGAGAGTCTACAGCTGATTTG 795
OY 41 LysLysGlnAlaIleGluAspLysGluAlaIleThrAlaIleGluAlaIleSerSerAsp 60
DB 796 ACAGCAGCAGCGGTACCGCTACTGTCGACAGCAGCGGCAAGCTGAAATGCTGGGCGACGA 855
OY 61 AlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80
DB 856 GCTTGGGAAGCAGCGGCA---GCAGCAGATGCTCTGACAAAGCCAAAGCAGATGCCCTT 912
OY 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
DB 913 AAGAAATTCACAAATATGAGTAGTACTATTACAGATCTATCAACATGCAATGCCAA 972
OY 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValAlaGlnSerAspAlaIle--- 114
DB 973 ACTGTTGAAGCCCTTAAAGACCTTCACAGCACAAGTTGTAATCAGCAGAAAGCGCGT 1032
OY 115 ---SerAspAlaLeuGlnSerGluGluAlaGluValAlaGlnSerAspAlaIle 127
DB 1033 ATTTCGAAAGCAACAGATGCTTATCTGATTTCTTAAATTCACAAACCTGCTGAGAT 1092
OY 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
DB 1093 ACTGTTAAATCAATGATTAATGATTAAGCTAAAGCTTAAAGCTTAAAGCAGAGA 1143

RESULT 16
AA247931
ID AA247931 standard; cDNA; 1576 BP.
XX
AC AA247931;
XX
DT 10-MAR-2000 (first entry)
XX
DE Streptococcus strain G 148 protein encoding cDNA.
XX
KW Streptococcus strain G 148; protein G'; protein G primer; liposome;
KW liposomal delivery complex; connecting moiety; antibody; Fc region;
KW diagnostic; therapeutic agent; antibiotic; antidepressant; antiviral;
KW antitumorigenic; cytokine; hormone; imaging agent; neurotransmitter;
KW stimulant; cytotoxic agent; malignant cell; ss.
XX
OS Streptococcus sp.
XX
FH Key Location/Qualifiers
FT CDS 292..603
FT /tag= a
FT CDS 705..1262
FT /tag= b
FT /product= "protein G'"
FT /trans_except= (pos:705..707,aa:Met)
XX
PN WO9959643-A2.
XX
PD 25-NOV-1999.
XX
PF 19-MAY-1999; 99MO-US11177.
XX
PR 20-MAY-1998; 98US-0086347.
XX
PA (SDGS-) SDG INC.
XX
PI Lau JR:
XX

```

```

DR WPI: 2000-062383/05.
DR P-PSDB: AA57610, AA57611.
XX
PT New liposome constructs comprising a liposome connected to an antibody,
XX used, e.g. for delivery of cytotoxic agents to malignant cells
XX
PS Disclosure: Fig 1; 32pp: English.
XX
XX
The present invention describes a construct for connecting an antibody
or antibody fragment to a liposome. The construct comprises protein G'
or a linking moiety for connecting the protein G' to the liposome.
CC Also described are: (1) a liposomal delivery vehicle comprising: (a) a
CC liposome; and (b) a connecting moiety connected to the liposome, which
CC specifically binds the Fc region of an antibody, for connecting the
CC antibody to the liposome; and (2) forming a liposomal carrier comprising:
CC (a) providing liposomal components having binding moieties, for forming
CC a liposome; (b) providing a construct comprising a linking moiety and a
CC connecting moiety bound together; and (c) combining the liposomal
CC components and the construct and sonicating the combination so that the
CC binding moieties are exposed on a surface of the liposome, for binding to
CC the linking moiety, and to facilitate the binding between them. The
CC products can be used for the delivery of diagnostic or therapeutic
CC agents. The liposomes may contain or may be associated with a diagnostic
CC or therapeutic agent, e.g. antibiotics, hormones, imaging agents,
CC antitumorigenic, antivirals, cytokines, hormones, hormones, imaging
CC neurotransmitters, or stimulants. They can be used particularly for the
CC delivery of cytotoxic agents to malignant cells. The protein G'
CC connecting moiety provides a liposomal delivery complex having improved
CC targeting efficiency. As a result of the binding between protein G' and
CC the Fc region of antibodies, protein G' shields the Fc regions of the
CC attached antibodies from non-specific binding to cell-surfaces, other
CC proteins, and anatomical structures. The present sequence encodes
CC an unidentified protein and protein G'.
XX
SQ Sequence 1576 BP; 572 A; 270 C; 330 G; 404 T; 0 other:

Alignment Scores:
Pred. No.: 2,216-07 Length: 1576
Score: 155.00 Matches: 81
Percent Similarity: 25.75% Conservative: 13
Best Local Similarity: 22.19% Mismatches: 62
Query Match: 19.95% Indels: 210
DB: 21 Gaps: 7

US-09-847-539A-6 (1-159) x AA247931 (1-1576)
OY 1 ValAspSerProIIleGIuInProAArgIleIleProAsnGlyGlyThrIleuThrAsnLeu 20
DB 391 GTTGACTCAGCAATGCAACATACCCCAATTATTCGTAATGCTGCTCAATTAATTAATCTTT 450
OY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
DB 451 CTGGGAATTCAGAGCAACACTGCTTGGCTAATGAGAGAGTCTACAGCTGATTTG 510
OY 41 LysLysGlnAlaIleGluAspLysGluAlaIleThrAlaIleGluAlaIleSerSerAsp 60
DB 511 ACAGCAGCAGCGGTACCGCTACTGTCGACAGCAGCGGCAAGCTGAAATGCTGGGCGACGA- 569
OY 61 AlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80
DB 570 GCTTGGGAAGCAGCGGCA---GCAGCAGATGCTCTGACAAAGCCAAAGCAGATGCCCTT 626
OY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeu 100
DB 627 AAGAAATTCACACTA-----GATGAAATTTAGCTGCTTACCTAAGACTGACACTTAC 680
OY 101 Gln----- 101
DB 681 AATTAATCCTTAATGTTAAACATTGAAGCGCAAAACAATACTGAAGCTGTGATGCT 740
OY 101 ----- 101
DB 741 GCTACTGCAAAAAGCTTCAACAAATAGGTTAAGCAGACAGCGTGTGACGCTGAATGG 800

```



```

QY 102 -----SerGluGluAlaGluValGln--- 109
Db 801 ACTTACAGCATGGCGACTAAGACCTTTACAGTTAGTGAATAACAGAAAGTGATGCG 860
QY 109 ----- 109
Db 861 TCTGATTAAACAGAGCGGTGACAACTTACAACTTGTATTATGTGTAACATTGAAA 920
QY 109 ----- 109
Db 921 GGCAGAACAACTACTGAAGCTGTGATGCTGCTACTGCGAGAAAAGCTTCAACAATAC 980
QY 110 ---SeraspAlaAlaSerAspAlaTrp----- 118
Db 981 GCTAACGACAGCGGTGTGACGCTGAATGAGCTTACGACGATGGACTAGACCTTTACA 1040
QY 118 ----- 118
Db 1041 GTTACTGAAAAACGAAAGTGTGATGCGTCTGAATTAAACACAGCGGTGACAACTTAC 1100
QY 118 ----- 118
Db 1101 AAACCTTTATTATGTTAAACATTGAAAGCGAACAACCTACTAAAGCAGTAGCGCA 1160
QY 119 -----GluLysAla----- 121
Db 1161 GAAACGACGAGAAAAGCCTTCAACAACTACGCAACAGCGTGTGATGCTTTGG 1220
QY 121 ----- 121
Db 1221 ACTTATGATGATGGAGCTAAGACCTTTACGTAAGTAAGTTACAGAGTTCCTGG 1280
QY 121 ----- 121
Db 1281 TGATGCACCACTGAACGAGAAAACGAGAACAGATCCCTGTTCCGTTAACTCC 1340
QY 122 -AlaThrProIleAlaLeuAspValLysLysThrLysAspThr----- 135
Db 1341 TGCACCTTCATTTGCTAAAGATGACGCTAAGAACAGCATCTACTAAGAAAGAGATGCTAA 1400
QY 136 -LysProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGly 155
Db 1401 AAACCAAGCACTAAGAAAGAGAGCGTAAAGAAAGCTGAACCTCTTCTACAACTGGTGA 1460
QY 155 uGluSerAsnPro 159
Db 1461 AGGAAGCAACCCA 1473

RESULT 17
AAN91099
ID AAN91099 standard; DNA; 798 BP.
AC AAN91099;
XX
XX 04-JUL-1990 (first entry)
XX
XX 'Protein G gene variant.
XX
XX Protein G; Immunoglobulin; Fc receptor; ds.
XX
XX Streptococcus sp.
XX
XX W08810306-A.
XX
XX 29-DEC-1988.
XX
XX 20-JUN-1988; 88WO-US02084.
XX
XX 19-JUN-1987; 87US-0063959.
XX
XX (GENE-) GENEX CORP.
XX

```

```

PI Fahnestock SR;
XX WPI; 1989-023848/03.
DR P-PSDB; AAP94785.
XX
XX Cloned protein G variant genes -
PT expressing proteins having immunoglobulin-binding properties of
PT protein G and derived from Streptococcus sp.
XX
XX Claim 4; Page 83; 116pp; English.
XX
XX Gene encodes protein G variant of non-pathogenic streptococcus allowing
CC isolation of the protein and variants, useful as bacterial Fc receptors
CC eg in purification and detection of Abs., screening of hybridoma clones
CC and treatment of disease.
XX
XX Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 5, 06e-06 Length: 798
Score: 139.00 Matches: 58
Percent Similarity: 35.65% Conservative: 24
Best Local Similarity: 25.22% Mismatches: 64
Query Match: 17.89% Indels: 84
DB: Gaps: 8

US-09-847-539a-6 (1-159) x AAN91099 (1-798)
QY 4 ProIleGluGlnProArgIleIleProAsnGlyLysThrLeuThrAsnLeuGlyAsn 23
Db 100 CCAATCGAGATATACCCCAATTATTCGTAATGCTGTGAATTAACCTTCTCTGGGAAT 159
QY 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGlu----- 39
Db 160 TCAGAGACACACCTGCTTGGCTTAATGAAGAGAGTCTACAGCTGGTACCCATTACT 219
QY 40 -----LeuLysLysGlnAlaIleGluAspLysGluAlaThr 51
Db 220 AAGACTGACACTTCAAAATTATTCCTTAATGCTTAAACATTGAAAGCGAACAACACTACT 279
QY 52 ThrAlaIleGluAlaAlaSerSerAsp-----Ala 61
Db 280 GAAGCTGTGATGCTGCTACTGACGAGAAAAGTCTTCAACAATACGCTAACGACACGGT 339
QY 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln-----SerGluGluAla 77
Db 340 GTTGACGGTGAATGACCTTACGACGATGCGACTAAGACCTTTACAGTTACTGAAAACCA 399
QY 78 AlaValValLysAlaAsp-----Asn 84
Db 400 GAAGTGAATCGATGCGCTGGAATTAAACACACCGCTGACAACTTACAAACCTTATTAT 459
QY 85 AlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu 104
Db 460 GGTAAACATTGAAGAGCGAACAACACTTAACAGATAGACGAGAAAGCTGCGAAGAAAA 519
QY 105 Ala---GluValValGlnSerAspAsnAlaIleSerAspAlaTrp----- 118
Db 520 GCCTTCAACAATACGCTAAGACGACACGCGTGTGATGCTGTGGACTTATGATGATGCG 579
QY 118 ----- 118
Db 580 ACTAAGACCTTTACGTAACGTGAATGTTACAGAGGTTCTCGACGATGATGACCAACT 639
QY 119 -----GluLys-----AlaAlaThrProIle 125
Db 640 GAACCAAGAAAACGAGAAAGCAAGTATCCCTTGTTCGTTAACTCCTGCAACTCCAAAT 699
QY 126 AlaLeuAspValLysLysThrLysAspThr-----LysProValVal 139
Db 700 GCTAAAGATGACGCTAAGAGAAAGACGATTAAGAAAGATGCTTAAAAACCGAAGCT 759
QY 140 LysLysGluGluArgGlnAsnValAsnThr 149

```

```

Db      760  AAGAAAGATGACGCTAAGAAAGCTGAACCT 789
      |||||:||||: |||
RESULT 18
ID      AAQ06009 standard; DNA; 798 BP.
XX
XX      AAQ06009;
AC
XX      17-JAN-1991 (first entry)
DT
XX      Sequence encoding Protein G variant.
DE
XX      Immunoglobulin; ds.
XX
XX      Streptococcus sp. Lancefield Group G strain.
OS
XX      US4956296-A.
PN
XX      11-SEP-1990.
PD
XX
XX      20-JUN-1988; 88US-0209236.
PF
XX
XX      20-JUN-1988; 88US-0209236.
PR
XX      14-FEB-1986; 86US-0829354.
PR
XX      23-APR-1986; 86US-0854887.
PR
XX      17-FEB-1987; 87MO-US00329.
PR
XX      19-JUN-1987; 87US-0063959.
XX
XX      (GENE-) GENEX CORP.
XX
XX      Fahnstock SR;
PI
XX
XX      WPI: 1990-297491/39.
DR
XX      F-PSDB; MAR07004.
PT
XX      Recombinant Protein G variants - obtd. using a cloned gene
PT      encoding Protein G from Streptococcus sp., used for blinding
XX      immunoglobulin.
XX
XX      Claim 4; Column 47; 48pp; English.
XX
XX      Sequence may be incorporated into a non-pathogenic host eg. E.coli,
CC      where they may be expressed at high levels. The proteins have a
CC      higher binding efficiency and capacity for immunoglobulin, and may
CC      be used for purifying, detecting and isolating antibodies.
XX
XX      Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;
SO
Alignment Scores:
Pred. No.: 5,06e-06 Length: 798
Score: 139.00 Matches: 58
Percent Similarity: 35.65% Conservative: 24
Best Local Similarity: 25.22% Mismatches: 64
Query Match: 17.89% Indels: 84
DB: 11 Gaps: 8
US-09-847-539A-6 (1-159) x AAQ06009 (1-798)
OY      4 ProIleGIuGIInProArGIleIleProAsnGIyGIYThIleuThraSleuLeuGIyAsn 23
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      100 CCAATGGAAGATACCCCAATTATTCGTAATGCTGTAATTAATTAATCTTCTGGGGAAT 159
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      24 AlaProGIuLysLeuAlaLeuArgAsnGIuGIuArgAlaIleAspGIu----- 39
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      160 TCAGAGACAACACTGGCTTGGCTTAATGAAGAGAGTGTCTACAGCTGGTAACCATTA 219
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      40 -----LeuLysIyGIuAlaIleGIuAspIySGIuAlaThr 51
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      220 AAGACTGACACTTCAAAATTATTCCTTAATGTGAANAACATTGAAGAAGCGAACAACACT 279
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      52 ThraIleGIuAlaIleAspSerAsp-----Ala 61
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db      280 GAAGCTGTTGATGCTGCTACTGACGAAAAAGTCTCAACAATACGCTAAACGACAAACGT 339
OY      62 LeuGIuAlaLeuAlaAspGIInThraSplaIleGIuIn-----SerGIuAla 77
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      340 GTTGACGCGTGAATGACTTACGACGATGCAGCTTAAGACCTTACAGTTACTGTAAAAACA 399
OY      78 AlaValAlaLysAlaAsp-----Asn 84
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      400 GAAGTATCATGATCGCTGCTCAATTACACCAAGCCGTGACAACTTACAACTTGTATTAA 459
OY      85 AlaIleSerAspAlaLeuGIuAlaLeuAlaAspGIInThraSplaIleGIuInSerGIu 104
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      460 GGTAAAAACATTGAAGGCGAACAACACTTAAGCAGTAGACCGCAACAACTGCAGAAAAA 519
OY      105 Ala-----GIuValAlaInSerAspAsnAlaIleSerAspAlaTrp----- 118
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      520 GCCTTCAACAAATACGCTTAAGCAGCAAGGTTGATGCTTGGACTTATGATGATGCG 579
OY      118 ----- 118
DB      580 ACTAAGACCTTTAGCGTAACTGAATGTTACAGAGTTCCTGAGTGATGACCAACT 639
OY      119 -----GIuLys-----AlaIleThrProIle 125
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      640 GAACAGAAAAAACAGACAGCAAGTATCCCTTGTTCGTTTACCTCCTCACTCCAAAT 699
OY      126 AlaLeuAspValLysIyThraSplaIleGIuIn-----LysProVal 139
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      700 GCTAAGATGACGCTAAGAAAGCGATTAAGCAAGAAGATGCTAAAAAACGAAAGCT 759
OY      140 LysIyGIuGIuArgGIuAsnValAsnThr 149
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      760 AAGAAAGATGACGCTAAGAAAGCTGAACCT 789
XX
RESULT 19
AAQ10007
ID      AAQ10007 standard; DNA; 798 BP.
XX
XX      AAQ10007;
AC
XX      13-MAR-1991 (first entry)
DT
XX
XX      Type 4 GX7809 protein G variant.
DE
XX
XX      Immunoglobulins; Ig; ds.
XX
XX      Streptococcus sp GX7809.
OS
XX      US497247-A.
PN
XX
XX      11-DEC-1990.
PD
XX
XX      19-MAY-1989; 89US-0354264.
PF
XX
XX      19-MAY-1989; 89US-0354264.
PR
XX      19-MAY-1989; 89US-0354264.
PR
XX      14-FEB-1986; 86US-0829354.
PR
XX      23-APR-1986; 86US-0854887.
PR
XX      17-FEB-1987; 87MO-US00329.
PR
XX      19-JUN-1987; 87US-0063959.
PR
XX      20-JUN-1988; 88US-0209236.
XX
XX      (GENE-) GENEX CORP.
XX
XX      Fahnstock SR, Lee T, Wroble MH;
PI
XX      WPI: 1991-006758/01.
DR
XX      P-PSDB; AAR10011.
PT
XX      Immobilised protein G variants - used for detection, isolation
XX      and purification. Immunoglobulin(s) and immunoglobulin fragments
XX
XX      Disclosure; Column 11; 52pp; English.
XX

```

CC Protein G variant product carries active regions B1 and B2. It may  
 CC be immobilised and exhibits different binding profiles. The bound  
 CC protein is useful in purification and detection of Igs and fragments.

XX Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

# Alignment Scores:

Pred. No.:	5,06e-06	Length:	798
Score:	139.00	Matches:	58
Percent Similarity:	35.65%	Conservative:	24
Best Local Similarity:	25.22%	Mismatches:	64
Query Match:	17.89%	Indels:	84
		Gaps:	8

US-09-847-539a-6 (1-159) x AA064648 (1-798)

```

QY  4  ProIleGIuGInProArGIlleIleProAsnGIyGIyThrLeuThrAsnLeuLeuGIyAsn 23
    |||||
Db 100 CCATCGAAGATACCCCAATTAATTCGTAATGCTGTAATTAATTAATTAATTCCTGCGGAAAT 159
QY  24  AlaProGIuLysLeuAlaLeuArGAsnGIuArGAlaIleAspGIu----- 39
    : : : |||||
Db 160 TCAGAGACACACTGCTGCTTGCCTAATGAAAGAGTGTCTACAGCTGGGTACCATTAAT 219
QY  40  -----LeuLysGIuAlaIleGIuAspLysGIuAlaIleThr 51
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 AAGACTGACACTTACAAATTAATTCCTTAATGTAACATTAATGAAAGCGGAACAACTACT 279
QY  52  ThrAlaIleGIuAlaIleAspSerAsp-----Ala 61
    |||||
Db 280 GAAGCTGTGATGCTGCTACTGACGAAAGTCTTCAACAAATACGCTAACGACACGGT 339
QY  62  LeuGIuAlaLeuAlaAspGIuInThrAspAlaLeuGIu-----SerGIuAlaIle 77
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 340 GTTACAGCTGAATGACCTTAACAGATGCGACTAAGACCTTACAGTTACTGAAACCA 399
QY  78  AlaValIleLysAlaAsp-----Asn 84
    |||||
Db 400 GAAGTGAATGCTGCTGCTGAATTAACACAGCCGCTGACAACTTCAAACTTGTATTAT 459
QY  85  AlaAlaSerAspAlaLeuAlaLeuAlaAspGIuInThrAspAlaLeuGIuInSerGIuGIu 104
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 460 GGTAAACATGAAAGCGAAGCAACTAATTAAGCAGTACAGCAGAACTGCAGAAATA 519
QY  105  Ala---GIuValIleGIuInSerAspAlaIleSerAspAlaIleP----- 118
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 520 GCCTTCAACAACTACGCTAACGACAGCGTGTGATGCTGTGACTTATGATGATGCG 579
QY  118  ----- 118
Db 580 ACTAAGACCTTTACGGTAATGAAAGGTTACAGAGTTCCTCGAGGTGATGACCAACT 639
QY  119  -----GluLys-----AlaAlaThrProIle 125
    |||||
Db 640 GAACCAAGAAACGAAAGCAAGATATCCCTCTGTTCCTTAACCTCCGCAACTCCAAAT 699
QY  126  AlaLeuAspValLysLysThrLysAspThr-----LysProValVal 139
    ||| ||| ||| |||||
Db 700 GCTAAAGATGACGCTAAGAAAGAGATTAAGAAAGATGCTAATAAAGCAAGAGCT 759
QY  140  LysLysGIuGIuArGAsnGIuArGAlaIleAspGIu----- 39
    |||||
Db 760 AAGAAAGATGACGCTAAGAAAGCTGAAACT 789

```

## RESULT 20

AA064648 standard; DNA: 798 BP.

AC AA064648;

DT 06-JAN-1995 (first entry)

XX Igc-binding streptococcus Protein G variant coding sequence.

XX Streptococcus Protein G; variant: Igc binding activity;  
 KW Immunoglobulin; Lancefield Group G; bacterial Fc receptor; ss.  
 XX Streptococcus sp.

XX	Key	Location/Qualifiers
XX	sig-peptide	1..90
FT	mat-peptide	/tag- a /note="encodes 30 amino acid secretion sequence" /tag- b /product= Protein_G_variant

US5312901-A.

```

XX 17-MAY-1994.
XX 14-FEB-1986; 86US-0829354.
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 19-JUN-1987; 87US-0063959.
PR 20-JUN-1988; 88US-0209236.
PR 19-JUN-1990; 90US-0540169.
PR 21-APR-1992; 92US-0871539.
PA (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX Fahnstock SR;
PI WPI: 1994-159179/19.
DR P-PSDB: AAR53294.
PT New recombinant streptococcal protein G variants - useful for
PT antibody detection and purification and for therapy
XX Disclosure; Columns 9-12 : 48pp; English.
XX Protein G isolated from Streptococcus has Igc-binding activity which
CC has been localised to the B repeating structure (see AAR53290).
CC The nucleotide sequence AA064648 codes for a claimed streptococcal
CC Protein G variant comprising the B domain active site and retaining
CC Igc-binding activity.
XX Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;
SQ

```

## Alignment Scores:

Pred. No.:	5,06e-06	Length:	798
Score:	139.00	Matches:	58
Percent Similarity:	35.65%	Conservative:	24
Best Local Similarity:	25.22%	Mismatches:	64
Query Match:	17.89%	Indels:	84
		Gaps:	8

US-09-847-539a-6 (1-159) x AA064648 (1-798)

```

QY  4  ProIleGIuGInProArGIlleIleProAsnGIyGIyThrLeuThrAsnLeuLeuGIyAsn 23
    |||||
Db 100 CCATCGAAGATACCCCAATTAATTCGTAATGCTGTAATTAATTAATTAATTCCTGCGGAAAT 159
QY  24  AlaProGIuLysLeuAlaLeuArGAsnGIuArGAlaIleAspGIu----- 39
    : : : |||||
Db 160 TCAGAGACACACTGCTGCTTGCCTAATGAAAGAGTGTCTACAGCTGGGTACCATTAAT 219
QY  40  -----LeuLysGIuAlaIleGIuAspLysGIuAlaIleThr 51
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 AAGACTGACACTTACAAATTAATTCCTTAATGTAACATTAATGAAAGCGGAACAACTACT 279
QY  52  ThrAlaIleGIuAlaIleAspSerAsp-----Ala 61
    |||||
Db 280 GAAGCTGTGATGCTGCTACTGACGAAAGTCTTCAACAAATACGCTAACGACACGGT 339

```

PR	06-MAY-1999	9905-01332486
PR	06-MAY-1999	9905-01332867
PR	07-MAY-1999	9905-01332883
PR	11-MAY-1999	9905-01342566
PR	14-MAY-1999	9905-01342428
PR	14-MAY-1999	9905-01342419
PR	14-MAY-1999	9905-01343221
PR	14-MAY-1999	9905-01343470
PR	18-MAY-1999	9905-01343478
PR	19-MAY-1999	9905-01349461
PR	20-MAY-1999	9905-01351524
PR	21-MAY-1999	9905-01355153
PR	24-MAY-1999	9905-01358529
PR	25-MAY-1999	9905-01360211
PR	27-MAY-1999	9905-01363692
PR	28-MAY-1999	9905-01367822
PR	01-JUN-1999	9905-01372722
PR	03-JUN-1999	9905-01375728
PR	04-JUN-1999	9905-01375702
PR	07-JUN-1999	9905-01377274
PR	08-JUN-1999	9905-01380094
PR	10-JUN-1999	9905-01385409
PR	10-JUN-1999	9905-01388477
PR	14-JUN-1999	9905-01391119
PR	16-JUN-1999	9905-01394552
PR	16-JUN-1999	9905-01394593
PR	17-JUN-1999	9905-01394532
PR	18-JUN-1999	9905-01394554
PR	18-JUN-1999	9905-01394455
PR	18-JUN-1999	9905-01394650
PR	18-JUN-1999	9905-01394669
PR	18-JUN-1999	9905-01394612
PR	18-JUN-1999	9905-01394557
PR	18-JUN-1999	9905-01394576
PR	18-JUN-1999	9905-01394558
PR	18-JUN-1999	9905-01394558
PR	18-JUN-1999	9905-01394592
PR	18-JUN-1999	9905-01394650
PR	21-JUN-1999	9905-01396117
PR	21-JUN-1999	9905-01396197
PR	22-JUN-1999	9905-01400353
PR	23-JUN-1999	9905-01400354
PR	23-JUN-1999	9905-01400354
PR	24-JUN-1999	9905-01400595
PR	24-JUN-1999	9905-01400595
PR	28-JUN-1999	9905-01400823
PR	29-JUN-1999	9905-01400921
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-01418442
PR	01-JUL-1999	9905-01421554
PR	02-JUL-1999	9905-01420055
PR	06-JUL-1999	9905-01423909
PR	08-JUL-1999	9905-01428083
PR	09-JUL-1999	9905-0142820
PR	12-JUL-1999	9905-01428277
PR	13-JUL-1999	9905-01433524
PR	14-JUL-1999	9905-01435324
PR	15-JUL-1999	9905-01440005
PR	15-JUL-1999	9905-01440005
PR	16-JUL-1999	9905-01440865
PR	19-JUL-1999	9905-01443433
PR	19-JUL-1999	9905-01443433
PR	19-JUL-1999	9905-01443433
PR	20-JUL-1999	9905-01443552
PR	20-JUL-1999	9905-01443633
PR	20-JUL-1999	9905-01446884
PR	21-JUL-1999	9905-01446814
PR	21-JUL-1999	9905-01450866
PR	21-JUL-1999	9905-01450868
PR	22-JUL-1999	9905-01450875
PR	22-JUL-1999	9905-01450887



FT	Key	Location/Qualifiers
FT	promoter	198..203
FT		/*tag- a
FT		/note- "putative promoter sequence"
FT	promoter	230..235
FT		/*tag- b
FT		/note- "putative promoter sequence"
FT	promoter	249..254
FT		/*tag- c
FT		/note- "putative promoter sequence"
FT	promoter	264..269
FT		/*tag- d
FT		/note- "putative promoter sequence"
FT	RBS	275..281
FT		/*tag- e
FT	CDS	288..1527
FT		/*tag- f
FT	s1g_peptide	288..389
FT		/*tag- g
FT	mat_peptide	390..1524
FT		/*tag- h
PN	W09507300-A.	
PD	16-MAR-1995.	
XX		
XX	06-SEP-1994;	94MO-SE00825.
PR	06-SEP-1993;	93SE-0002856.
XX		
PA	(GUS5/) GUS5 B.	
PA	(JONS5/) JONSSON H.	
PA	(LIND/) LINDBERG M.	
PI	Gus5 B, Jonsson H, Lindberg M;	
XX		
DR	WPI; 1995-123386/16.	
DR	P-PSDB; AAR71670.	
XX		
PT	DNA encoding a serum albumin binding protein - used to obtain	
PT	prods. for sepsis, detection, quantification, protein	
PT	stabilisation or vaccine development	
XX		
PS	Disclosure; Page 29-30; 38pp; English.	
XX		
CC	A phage lambda EMB3 library of S. dysgalactiae 8215 DNA was	
CC	analyzed for alpha-2-macroglobulin-, BSA- and IgG-binding activity.	
CC	Clone lambda-SD1, which expressed all 3 activities, was analyzed to	
CC	obtain DNA encoding the SA-binding protein, MAG. The mag gene	
CC	(given in Q86080) encoded a 44 kDa MAG protein (R71670).	
CC	Recombinant, immobilized MAG was used for serum albumin affinity	
CC	purification, detection and assay.	
XX		
SO	Sequence 1555 BP; 582 A; 242 C; 305 G; 426 T; 0 other;	
	Alignment Scores:	
	Pred. No.:	0.00489
	Score:	114.50
	Percent Similarity:	34.22%
	Best Local Similarity:	26.20%
	Query Match:	14.74%
	DB:	16
	US-09-847-539A-6 (1-159) x AAO86080 (1-1555)	
OY	36 AAlAlleaspgluenuylsYsGlnAlAlleGluASP-	47
DB	903 GCTATTGAAATTCGTGAGAAATACGCAATTCGCGATTACATATTAATTAATTAAT	962
OY	48 - - - - -LysGluAlaTrpTrpAlaAlleGlu- - - - -AlaAlaseiserAspAla	61
DB	963 GGTAAACATCGAAGGTGTGATCGCTCTTAAGATGAATTTTACTTCAAAACACAGA	1022

Oy	62	leunglualaleualaaspglnthrspalaleuglnserglucglnlualalevalValys	81
Db	1023	GTGATTGACGCACCTGATTAACCAACCAGCTTGTGACAACACTACAACTGGTTATC-----	1076
Oy	82	AlaaspasnalaalaseraspalaleugluAlaleualaaspglnthrspalaleugln	101
Db	1077	-----AMTGATAAACACTTGAAGAGCGAACAACACTACTAAAGCACTAGACCCACAAACT	1130
Oy	102	SerGlucglnuaL---gluvalvalgluseraspasnalalaseraspalatrpglu--	119
Db	1131	CGAGAAAAGAGCCCTCAAACAACATACGCTAACGAAACGCGTTGATGATGTTGGACTTAC	1190
Oy	119	-----	119
Db	1191	GATGATCGGACTAAGACCTTTACTGTAACTGAAGAATGTTACTGAAGTTCGGTGATGCA	1256
Oy	120	-----LysAlaAlathr	123
Db	1231	CCAAGTGAACAAAAAACCAAGCAAGTATCCCTCTTGTTCGGTTAACTCCTCAACT	1310
Oy	124	ProfilalealeuasparValLysLysThrLysAspThr-----	135
Db	1311	CCAATTGCTAAAGATGACCGCTAAGAAAGACGATACTAAGAAGACGATCTAAGAAAGAA	1370
Oy	136	-----LysProvalValLysLysgluIuarlglnshsnValasnthreuprothr	152
Db	1371	GATGCTAAAAAACCGAAGCTAAGAAAGAAAGAGTAGTAAAGAAAGCTGCACACTTCTACA	1430
Oy	153	ThrglygluIuserasnpro	159
Db	1431	ACTGTTGAAGAACCAACCCA	1451
<hr/>			
RESULT 23			
AAQ89197	ID	AAQ89197 standard; DNA; 1555 BP.	
XX	AC	AAQ89197;	
XX	DT	22-SEP-1995 (first entry)	
XX	DE	S. dysgalactiae mag gene.	
XX	KM	Mag; fast alpha-2-macroglobulin binding protein; FAM:	
XX	KW	plasma proteinase-inhibitor binding protein; ss.	
OS	XX		
XX	OS	Streptococcus dysgalactiae strain 8215.	
FH	FH	Location/Qualifiers	
FT	FT	promoter	
FT	FT	/note= "a	
FT	FT	/tag= "putative promoter region"	
FT	FT	230..235	
FT	FT	/tag= b	
FT	FT	/note= "putative promoter region"	
FT	FT	249..254	
FT	FT	/tag= c	
FT	FT	/note= "putative promoter region"	
FT	FT	264..269	
FT	FT	/tag= d	
FT	FT	/note= "putative promoter region"	
FT	FT	276..281	
FT	FT	/tag= e	
FT	FT	CDS	
FT	FT	288..1527	
FT	FT	/tag= f	
FT	FT	sig_peptide	
FT	FT	288..389	
FT	FT	/tag= g	
FT	FT	mat_peptide	
FT	FT	390..1524	
FT	FT	/tag= h	
PX	PX	WO9507296-A.	
XX	PD	16-MAR-1995.	



PA (GUS5/) GUS5 B M.  
 XX Guss BM, Lindberg KM, Flock JI, Uhlen CEM;  
 XX WPI: 1987-277686/39.  
 DR P-PSDB: AAF70468.  
 XX  
 PT New recombinant DNA molecules - for producing proteins with  
 PT IgG-binding specificity of protein G or proteins A and G  
 XX  
 PS Example: Fig 2; 39pp: English.  
 CC A recombinant DNA molecule containing a nucleotide sequence which  
 CC codes for a protein or polypeptide having the same IgG specificity as  
 CC protein G from *Streptococcus G148* (AAN70757) is claimed. See, for  
 CC example, AAN70754, AAN70755 and AAN70756.  
 XX  
 SQ Sequence 1469 BP; 545 A; 280 C; 297 G; 347 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.00847 Length: 1469  
 Score: 112.00 Matches: 44  
 Percent Similarity: 39.44% Conservative: 12  
 Best Local Similarity: 30.99% Mismatches: 62  
 Query Match: 14.41% Indels: 24  
 DB: Gaps: 5  
 US-09-847-539A-6 (1-159) x AAN70757 (1-1469)  
 QY 41 LysLysGlnAlaIleGlnAspLysGlnAlaIleThrAlaIleGlu-----Ala 56  
 DB 943 AAACCGAAGTGCATGCTGTAATTACACACGCGCTGCAACTTACAACCTTGT 1002  
 QY 57 AlaSerSerAspAlaLeu-----GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 74  
 DB 1003 ATTAAATGTTAAACATTGGAAGCCGAAACACTACTTAAAGCACTAGACGAGAACTGCA 1062  
 QY 75 GluGlnAlaIleValValLysAlaAspAlaIleSerAspAlaLeuGlnAlaLeuAla 94  
 DB 1063 GAAAAAGCCTTCAACATATACGCTAACGACAAAGCGTGTGATGCTTGGACTTATGAT 1122  
 QY 95 AspGlnThrAspAlaLeuGlnSerGlnGlnAlaIleValValGlnSerAspAlaIleAla 114  
 DB 1123 GATGGCACTAAGACCTTACGCTTACGTAATG---GTTACAGAGGCTCTCGTGATGCA 1179  
 QY 115 SerAspAlaIleThrGlnLys-----AlaAlaIleThr 123  
 DB 1180 CCAACTGAACCAAGAAAAACAGACAGCAATCCTCTTCTTCCGTTAACTCCTGCAACT 1239  
 QY 124 ProIleAlaLeuAspValLysLysThrLysAspThr-----LysPro 137  
 DB 1240 CCAATTGCTAAGATGATGCTTAAAGAAAGACATACAGAAAGAGATGCTTAAAGAACCA 1299  
 QY 138 ValValLysLysGlnGlnAlaIleGlnSerValAsnThrLeuProThrThrLysGlnGlnSer 157  
 DB 1300 GAAGCTTAAGAAAGAGACGCTAAGAAAGCTGAAGAACTCTTCTTCAACTGCTGAAGGACGC 1359  
 QY 158 AsnPro 159  
 DB 1360 AACCCA 1365  
 RESULT 25  
 ID ABL20337 standard: DNA; 2121 BP.  
 AC ABL20337;  
 DT 26-MAR-2002 (first entry)  
 XX Drosophila melanogaster genomic polynucleotide seq ID NO 12484.  
 XX Drosophila: developmental biology; cell signalling; insecticide;  
 KW Pharmaceutical: gene; ds.

XX OS Drosophila melanogaster.  
 XX WO200171042-A2.  
 XX  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NT.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 DR WPI: 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Claim 1; SEQ ID NO 12484; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
 CC sequences (AB057737-AB072072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp://wipo.int/pub/publ/pdb-pct-sequences.  
 XX  
 SQ Sequence 2121 BP; 533 A; 638 C; 587 G; 363 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.0169 Length: 2121  
 Score: 111.00 Matches: 41  
 Percent Similarity: 46.92% Conservative: 20  
 Best Local Similarity: 31.54% Mismatches: 43  
 Query Match: 14.29% Indels: 26  
 DB: Gaps: 6  
 US-09-847-539A-6 (1-159) x ABL20337 (1-2121)  
 QY 4 ProIleGlnProAlaGlnIleProAsnGlyGlyThrLeuThrAsnLeuGlnLysAsn 23  
 DB 1150 CCCCTGACGATCCCTCGATG-----GAG 1176  
 QY 24 AlaProGlnLysLeuAlaLeuArgAsnGlu-----GluArgAlaIle 37  
 DB 1177 GCCAAGGAAACACAGCGTGTGCGCAGAGCGCTTCTTACCGTTTACCAAGAGCGCTCTC 1236  
 QY 38 AspGlnLeuLysLysGlnAlaIleGlnAspLysGlnAlaIleThrThrAlaIleGlnAlaIle 57  
 DB 1237 GACGAGGTGAC-----TTGCAAAAGAACCGTCCAGAAAGCCATGAAGCAAG 1287  
 QY 58 SerSerAspAlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeu-----GlnSerGlu 75  
 DB 1288 CGTAGCGAAGAG---AGGCTCTCAAGAGAGAGAGATCGCTGAGCGCTGAGCCGAA 1344  
 QY 76 GluAlaIleValValLysAlaAspAlaIleSerAspAlaLeuGlnAlaLeuAlaAsp 95  
 DB 1345 AAGCAGCGTGTCTAAGCAAAAGCTGCGAGAGAGCGCCCAAAATCGCTGCCGAA 1404  
 QY 96 Gln-----ThrAspAlaLeuGlnSerGlnGlnAlaIleValValGlnSerAspAsn 112  
 DB 1405 GAGGCTCTCTGCTGCAAGCTGCCCCCAAAAGCAGCGAAAGAGCTTAAAGCCCTTAAA 1464  
 QY 113 AlaAlaSerAspAlaIleThrGlnLysAlaIle 122





PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0144816.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151348.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154479.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.:	0.00765	Length:	812
Score:	109.50	Matches:	41
Percent Similarity:	43.68%	Conservative:	35
Best Local Similarity:	23.56%	Mismatches:	65
Query Match:	14.09%	Indels:	33
DB:	21	Gaps:	6

US-09-847-539a-6 (1-159) x AAC38790 (1-812)

QY 7 GlnProAgtle-----lLeProAsnGlyGlyThrLeuThAsnLeu 20  
DB 87 CACCCCAAAATCACAAAAAATTTATATATATGCTGGC----- 131  
QY 21 LeuGlnAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaLe----- 37  
DB 132 TGTCGAGTAGACCAAGAAATCATACATGCAAAAGAGGCTGTTCTTCAACAG 191  
QY 38 -----AspGluLeuLysLys 42  
DB 192 CCTGCTCTGTAAGTAGTACAGAGGAAACAACAGAAAGCGCTGACAAACAGAAC 251  
QY 43 GlnAlaIleGluAspLysGluAlaThrThAlaIleGluAlaIleSerSerAspAlaLeu 62  
DB 252 CAGACAGTGGCTGAGACACAGTGAACAACTTCTGCGAAGCTAAGACATCTCCGCTC 311  
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleValLysAla 82  
DB 312 GAGCCGACCAAGGAAACACTCTGCTGCTGACCCAGAGTTCGACGAGAGTCT 371  
QY 83 AspaAlaIleAspAlaLeuGlnAlaLeuAlaIleAspGlnThrAspAlaLeuGln 101  
DB 372 TCTTCTGCTGCTGCTGAGGAGCGACGACATTGCACTGAGAAAGTGAAGAAGCTCCA 431  
QY 102 SerGluGluAlaGluAlaValGlnSerAspAsnAlaIleSerAspAlaIleTrpGluLysAla 121  
DB 432 ACAGAAAAAGCTGAA---GCAAAAGTGGAGCGGTTGCCCTAGCTGACCCGAAAG--- 485  
QY 122 AlaThrProIleAlaLeuAspValLysThrLysAspThrLysProValLysLys 141  
DB 486 ---GTTGAGTTCGACGAGGAGCGGAGGAAAGCGCTGAGAGTGAAGCTGAAGAGTTAG 542  
QY 142 GluGluArgGlnAsnValAsnThrLeuProThrThrGlyGlu 155  
DB 543 GCTGACCAAGTAAAGCTGAGGCTGACCACTAAAGCTGAG 584  
RESULT 28  
AAC46811  
ID AAC46811 standard; DNA; 1128 BP.  
XX  
XX AAC46811;  
AC  
AC 18-OCT-2000 (first entry)  
DT  
DT 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 51490.  
DE  
DE 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 51490.  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS  
OS Arabidopsis thaliana.  
XX

PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139464.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139839.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.

```

PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145085.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 0.0129
Score: 109.00
Percent Similarity: 39.20%
Best Local Similarity: 27.27%
Query Match: 14.03%
DB: 21
Gaps: 7

US-09-847-539a-6 (1-159) x AAC46811 (1-1128)
OY 27 LysLeuAlaLeuArgAsnGluIuArgAlaLeuAspGluLeuLys----- 42
||||| |||:||||| |||:||||| |||:|||||
DB 571 AAGCTGACCTTCANAAAGCTGGAGAAAGCATTTCAGACCTTAACGATTGAGCAAGA 630
OY 43 -----GlnAlaIleGluAsp-----LysGluAlaThrThrAla 53
||||| |||:||||| |||:||||| |||:|||||
DB 631 GAGAGGATCATCAACCATTTGAGATTCGAGAAAGCAAAAGATGCAAAACCMA 690
OY 54 IleglAlaIleSerSerSpAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 73
::: |||:||||| |||:||||| |||:|||||
DB 691 CTTCTGCTGTAGAACAGACGTCAGATTCGCTTTTTCAGACTTCAGACCAAAAGAT 750
OY 74 SerGluGlu----- 76
:::|||||
DB 751 GCAGAGCAACACGCCAGGACCGCTAAATGTTCAGATTTCGATTCAGATCA 810
OY 77 AlaAlaValAlaLysAlaAspAsnAla-----AlaSerAspAlaLeuGluAla 92
||||| |||:||||| |||:||||| |||:|||||
DB 811 TTACGCTGTGTGAAGAAAGTGGCTGTGAAGAAATTACGACAGCAACACCTTAAAGAGCT 870
OY 93 LeuAlaAspGlnThr--AspAlaLeuGlnSerGluAlaGluAlaValGlnSerAsp 111
|||:||||| |||:||||| |||:|||||

```

Db 871 TGCAGAGATGAACTCAACATTAGAGATGAGAGAGAAATTGAGATATATAAA 930  
 QY 112 AsnAlaIaSerAspAlaTpgIuLysAla-----AlaThrProIleAlaLeu 127  
 Db 931 GCTGCACACACAGATCTTTGAAGAAAGCTGAGATGGCTCAGACAGCTACAAATTGTGGT 990  
 QY 128 AspValLysLys----- 131  
 Db 991 GAGCTTGAGCTGAAAGAGCGCGTAAAGCAGCAGTACGATCTTGGCAGATCCAAATG 1050  
 QY 132 ---ThLysAspThrLysProValValLysLysGluGluArgGlnAsn 146  
 Db 1051 TGTGCCAAATCTACCAAGAGGTGTCTCAAGTCAAAACCAAGATCATCG 1098  
 RESULT 29  
 ABL11003  
 ID ABL11003 standard; cDNA; 4086 BP.  
 AC ABL11003;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27491.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR P-PsDB; ABB66900.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1: SEQ ID NO 27491; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AB857737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from Wipo  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SO Sequence 4086 BP; 1311 A; 960 C; 1056 G; 759 T; 0 other;

US-09-847-539a-6 (1-159) x ABL11003 (1-4086)  
 QY 11 IleProAsnGlyLysThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeuAlaLeu 30  
 Db 2796 GTCCCAAT-----ATTAATGTCAGGCCGCAAAATGATGTG 2834  
 QY 31 ArgAsnGlu-----GluArgAlaIleAspGluLeuLysLysGlnAla 44  
 Db 2835 GACTCAACAGCGCGGAGAGAGAGAGCTTCGACAGCCGAGTCCAGATCAAGAAAGAGCT 2894  
 QY 45 IleGluAspLysGluAlaThrThrAlaIleGluAlaIaSer-----SerAsp 60  
 Db 2895 GAGCTTGAAACATCAAGAGAAATTTATCAAGAGATTCACCAATTCCGATTCGAT 2954  
 QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValAla 80  
 Db 2955 TCCGTGCGGAATTAAGAGAGCGGTAACTGCTCTGAGGCGAGATGTGTCATCTA 3014  
 QY 81 LysAlaAspAsnAlaIaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
 Db 3015 GAGCCCGATACATACCCAAAGAGCTCTGATGATTAATTCGCCGAGCCGAGAAA--- 3071  
 QY 101 GlnSerGluGluAlaGluAlaValGlnSer---AspAsnAlaIaSerAspAlaTpgIu 119  
 Db 3072 CCTGATCAGAGAGAGACATTTGTTCAATCTGAAGAGATCGAGACAGAGCGGTTAGAT 3131  
 QY 120 LysAlaIaIaThr-----ProIleAlaLeuAspValLysLysThr 132  
 Db 3132 AGTCAGTAAAGCAGAGAGATGACCTGCTCCACCTGACGAGTACTACCAAAATG 3191  
 QY 133 LysAspThrLysProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThr 152  
 Db 3192 GAGATTGATGACACAGAGCGGAGAAAGCAGCCGAAATTAATGACTACTAGACGCGA 3251  
 QY 153 ThrGlyGluGlnSer 157  
 Db 3252 GACGAGAAAGAGAGCC 3266  
 RESULT 30  
 ABL11002/c  
 ID ABL11002 standard; cDNA; 6782 BP.  
 XX  
 AC ABL11002;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27488.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR P-PsDB; ABB66899.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX

## Alignment Scores:

Pred. No.: 0.0789  
 Score: 108.00  
 Percent Similarity: 47.27%  
 Best Local Similarity: 21.82%  
 Query Match: 13.90%  
 DB: 23

## Length:

Matches: 4086  
 Conservative: 42  
 Mismatches: 61  
 Indels: 26  
 Gaps: 6

PS Claim 1: SEQ ID NO 27488; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB857737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [ftp://wipo.int/pub/published\\_pcl\\_sequences](http://wipo.int/pub/published_pcl_sequences).  
 XX  
 XX Sequence 6782 BP; 1412 A; 1688 C; 1602 G; 2080 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 0.146 Length: 6782  
 Score: 108.00 Matches: 36  
 Percent Similarity: 47.27% Conservative: 42  
 Best Local Similarity: 21.82% Mismatches: 61  
 Query Match: 13.90% Indels: 26  
 DB: Gaps: 6  
 US-09-847-539a-6 (1-159) x AB111002 (1-6782)  
 QY 11 ILEPROASNGLYGlyThrLeuThraSLeuLeuGlyAsnAlaProGluLysLeuAlaLeu 30  
 DB 2291 GTCCCAAT-----ATAAATACGCACCCGAGAAAATGATGTCG 2253  
 QY 31 ArgAsnGlu-----GluArgAlaIleAspGluLeuLysGlnAla 44  
 DB 2252 GACTCAGAAAGCGCGAGCAGCAAGCCTTCGAAAGCCGACGATCAAGAAAGAGCGCT 2193  
 QY 45 ILEGLuAspLysGluAlaThrThraIleGluAlaIleAsp-----SerAsp 60  
 DB 2192 GAGCTTGAAACGATCAAGAGAAATTTATCAAGAGATTCCCAATTCGCGATTCCGAT 2133  
 QY 61 AlalaGluAlaIleuAlaAspGlnThraAspAlaLeuGlnSerGluGluAlaIleVal 80  
 DB 2132 TCCGTCGCGGATTTAAGAGAGCGGTACTGCTCGAGCGGAGAGATGATGTCATCTA 2073  
 QY 81 LysAlaAspAsnAlaIleAspAlaLeuGluAlaIleuAlaAspGlnThraAspAlaLeu 100  
 DB 2072 GAGCGCGATTAACATACGCAAGAGCGCTAGATGATTGATTCGCGAGCGCAAGAA--- 2016  
 QY 101 GlnSerGluGluAlaGluValAlaGlnSer---AspAsnAlaIleSerSerAspAlaTrpGlu 119  
 DB 2015 CCTGATTCAGAGAGACATTTGTCAATCTGAAGAAATGCGAGCAAGAGCGCTTAGAT 1956  
 QY 120 LysAlaAlaThr-----ProIleAlaLeuAspValLysLysThr 132  
 DB 1955 AGCTCAGTACGCGAGCAGATACCTGCTGCACCTACCGCATCTATCCCAACAAAG 1896  
 QY 133 LysAspThrLysProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThr 152  
 DB 1895 GAGATTGATGACACGACGAGCAAGCAAGCAAGCAAAATATGAGATCTGAGACGGCA 1836  
 QY 153 ThrGlyGluGlnSer 157  
 DB 1835 GACGAGAAAGAAAGC 1821  
 RESULT 31  
 AAX13883/C  
 ID AAX13883 standard; DNA; 486 BP.  
 XX  
 AC AAX13883;  
 XX  
 DT 19-MAR-1999 (first entry)  
 XX  
 DE Enterococcus faecalis genome contig SEQ ID NO:946.  
 XX  
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.  
 XX  
 OS Enterococcus faecalis.  
 XX  
 PN WO9850555-A2.  
 XX  
 PD 12-NOV-1998.  
 XX  
 PF 04-MAY-1998; 98MO-US08985.  
 XX  
 PR 14-NOV-1997; 97US-0066009.  
 PR 06-MAY-1997; 97US-0044031.  
 PR 16-MAY-1997; 97US-0046555.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Barash SC, Dillon PJ, Kunsch CA;  
 XX  
 PI WPI; 1999-045171/04.  
 XX  
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
 PT - used to develop products for the detection of Enterococcus and for  
 PT use in vaccines for prevention or attenuation of Enterococcus  
 PT infection.  
 XX  
 PS Claim 1: Page 2060; 2084pp; English.  
 XX  
 CC A computer readable medium has been developed which has recorded on it  
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
 CC AAX12938 to AAX1919 represent these nucleotide sequences which are  
 CC primary nucleotide sequences, also known as contigs. The computer-based  
 CC system can identify fragments of the Enterococcus faecalis genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of Enterococcus faecalis in samples. They can also be used for  
 CC diagnosing Enterococcal infection in an animal and monitoring  
 CC progression of disease, and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
 CC another related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.  
 XX  
 SQ Sequence 486 BP; 71 A; 110 C; 98 G; 205 T; 2 other;  
 Alignment Scores:  
 Pred. No.: 0.00673 Length: 486  
 Score: 107.50 Matches: 37  
 Percent Similarity: 47.20% Conservative: 22  
 Best Local Similarity: 29.60% Mismatches: 61  
 Query Match: 13.84% Indels: 5  
 DB: Gaps: 4  
 US-09-847-539a-6 (1-159) x AAX13883 (1-486)  
 QY 25 ProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysGlnAla 44  
 DB 478 CCAGCAGTGAAGCTACAGCAACGAGCAAGCATTAATCGAA---AAACAGCAAGCA 422  
 QY 45 ILEGLuAspLysGluAlaThrThraIleGluAlaIleAspSerSerAspAlaLeuGluAla 64  
 DB 421 GTACAGAGAAACCAATTCGATCAAAACCAAGTGTGACACTCGGAAAAA 362  
 QY 65 LeuAlaAspGlnThraAspAlaLeuGlnSerGluGluAlaIleValLysAlaAspAsn 84  
 DB 361 GAAAAAGACCATGATTAATCTGTTAAAGACCAACAGAGTGTCATCAAAACAA 302  
 QY 85 AlaIleAsp---AlalaGluAlaLeuAlaAspGlnThraAspAlaLeuGlnSerGlu 103  
 DB 301 GGTGATTCGATCAAGTACACAGACAGTACGACAA---CAAGCGGTCTAGACGAA 245  
 QY 104 GluAlaGluValAlaGlnSerAspAsnAlaIleAspAlaTrpGluLysAlaIleThr 123  
 DB 244 GCAAAAAAGTTGTG-----GATGAGCAACACCTTCGACGCTTTGAAAAAGCAAAAC 191

QY 124 Proilealeuaspvalylslysthrlyaspthrlyspvalylslysglu 143  
 Db 190 CAAGTGGCTACGATACCAAGCTGTGATGACCAACAAAAGTAGTCAGCTCAG 131  
 OY 144 ArgGlnasnValasn 148  
 Db 130 CAGACGTTAACCAAC 116

RESULT 32  
 AAT91293  
 ID AAT91293 standard; DNA; 989 BP.  
 XX  
 AC AAT91293;  
 XX  
 DT 27-APR-1998 (first entry)  
 XX  
 DE Rabbit enteropathogenic E. coli (RDEC-1) espB gene.  
 XX  
 KW EspB; rabbit enteropathogenic E. coli; RDEC-1; virulence factor;  
 KW pathogen; infection; diagnosis; therapy; vaccine; gastroenteritis;  
 KW ss.  
 XX  
 OS Escherichia coli.  
 XX

FM Key Location/Qualifiers  
 FT RBS 19..21  
 FT CDS /\*tag- a  
 FT 31..975  
 FT /\*tag- b

PN W09740063-A2.  
 XX  
 PD 30-OCT-1997.  
 XX  
 PF 23-APR-1997; 97MO-CA00265.  
 XX  
 PR 23-APR-1996; 96US-0015999.  
 XX  
 PA (UYBR-) UNIV BRITISH COLUMBIA.  
 XX  
 PI Finlay BB, Kenny B, Stein M;  
 XX  
 DR WPI: 1997-535772/49.  
 DR P-PSDB: AAM27286.  
 XX  
 PT EspA from entero-pathogenic or entero-haemorrhagic E. coli - used to  
 PT immunise against, and treat diseases caused by EspA producing  
 PT organisms  
 XX  
 PS Example 6; Fig 4B; 62pp; English.  
 XX  
 CC This sequence comprises the espB gene from a rabbit  
 CC enteropathogenic Escherichia coli (RDEC-1) strain. It codes for a  
 CC claimed EspB virulence factor (see AAM27286). The espA and espA  
 CC (see AAT91292) genes were obtained by PCR from RDEC-1 chromosomal  
 CC DNA using primers derived from the published sequence of  
 CC enteropathogenic E. coli. Claimed novel, isolated EspA polypeptides  
 CC are characterised by: (a) being a secreted protein from  
 CC enteropathogenic (EPEC) or enterohaemorrhagic (EHEC) E. coli;  
 CC and (b) having a mol. wt. of about 25 kD as determined by SDS-PAGE.  
 CC Isolated espA nucleic acids can be used in the recombinant  
 CC production of EspA polypeptides, and as probes in the diagnosis of  
 CC disease caused by EspA-producing E. coli. A claimed method of  
 CC producing a non-pathogenic E. coli comprising inserting a  
 CC selectable marker nucleic acid into the espA gene. EspA  
 CC polypeptides can be used to immunise a host susceptible to disease  
 CC caused by EspA-producing E. coli, and in a method for ameliorating  
 CC disease caused by EspA-producing organisms.  
 XX  
 SQ Sequence 989 BP; 278 A; 211 C; 240 G; 260 T; 0 other;

Alignment Scores:

Prod. No.: 0.0262 Length: 989  
 Score: 105.50 Matches: 45  
 Percent Similarity: 44.858 Conservative: 29  
 Best Local Similarity: 27.278 Mismatches: 72  
 Query Match: 13.58% Indels: 20  
 DB: 18 Gaps: 6

US-09-847-539A-6 (1-159) x AAT91293 (1-989)

OY 4 ProilegluInProarglleIleProasnGlyGlyThrLeuthrAsnLeuLeuGlyAsn 23  
 Db 305 CCGCTGCACAGCCGCT-CTGGTT-----GTTGGGCTATTTCATCATGATGGGGATC 357  
 OY 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysGlu 43  
 Db 358 TTAGGCTCTTTTTCGAGCAATTAACAGTCTACGAAAGCCGAGTGAT----- 405  
 OY 44 AlaIleGluAspLysGluAlaThrThr-----AlaIleGluAlaIleSerSerAsp 60  
 Db 406 ---ATTGCTCAAAAACCGCCTCTACATCTTCTAAGGCTATTGATGGCTTCTGATACT 462  
 OY 61 AlaLeuGluAlaLeuAlaAspGlnThr-----AspAlaLeuGlnSerGluGlu 76  
 Db 463 GCGACTAAACGTTGACTAGCGACGCAAGCGTTGCTGATGCTGTGAAGATGCATCC 522  
 OY 77 AlaAlaValAlaLysAlaAspAsnAlaIleSerAspAlaLeuAlaLeuAlaAspGln 96  
 Db 523 AGCGTGATGCGACAGGAGTACGACTACGACGACGAGAGCGCGCTACATCCGAGT 582  
 OY 97 ThrAspAlaLeuGlnSerGluGluAlaGluAlaValGlnSerAspAsnAlaIleSerAsp 116  
 Db 583 GGTGATGACATGCGCATTTTGTCTCAGACAGCTTCTAGCTGCTGAAACGCTGCAGAT 642  
 OY 117 AlaTrpGluLysAlaIleThrPro-----IleAlaLeuaspValylslysthr 132  
 Db 643 GCCGTCAGAGGAGGAGTCCGCGACGCCGCTTATGCGTCACTAGAT---AAGATTACT 699  
 OY 133 LysAspThrlyspvalylslysgluGluArgGlnAsnValAsnThrLeuProThr 152  
 Db 700 GGCTCTACACCATTTATGCGCTTACCATGTTGCCGACGACGAGACGAGACATTGCCACA 759

QY 153 ThrGlyGluLysSer 157  
 Db 760 ACCGTATCTGATCA 774

RESULT 33  
 AAQ84886  
 ID AAQ84886 standard; cDNA; 627 BP.  
 XX  
 AC AAQ84886;  
 XX  
 DT 11-OCT-1995 (first entry)  
 XX  
 DE Synthetic protein G gene, MD1.  
 KW Synthetic; protein G; IgG; Fc receptors; heavy chain; leukaemia;  
 KW lymphoma; cancer; autoimmune disease; ds.  
 OS Synthetic.  
 XX  
 PN W09506125-A.  
 XX  
 PD 02-MAR-1995.  
 XX  
 PF 23-AUG-1994; 94MO-US09141.  
 XX  
 PR 23-AUG-1993; 93US-0110653.  
 XX  
 PA (IMMU-) APPLIED IMMUNE SCI INC.  
 XX  
 PI Lee YM, Okarma TB, Talib S;  
 XX  
 XX WPI: 1995-106854/14.









AC AAS52179;  
XX  
DE 13-FEB-2002 (first entry)  
XX  
DE Staphylococcus aureus DNA for cellular proliferation protein #596.  
XX  
XX Antisense; ds; prokaryotic cellular proliferation gene;  
KM antibiotic; antibacterial; drug design.  
XX  
OS Staphylococcus aureus.  
XX  
PN WO200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US09180.  
XX  
XX 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
DR MPI: 2001-611495/70.  
DR P-PSDB: AAU34320.  
XX  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX Claim 27; Seq ID NO 4761; 511pp; English.  
XX  
CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 7434 BP; 3197 A; 1326 C; 1298 G; 1613 T; 0 other;

Alignment Scores:  
Pred. No.: 0.439 Length: 7434  
Score: 104.00 Matches: 51  
Percent Similarity: 37.80% Conservative: 28  
Best Local Similarity: 24.40% Mismatches: 70  
Query Match: 13.38% Indels: 60  
DB: 23 Gaps: 8

US-09-847-539a-6 (1-159) x AAS52179 (1-7434)

OY 1 ValAspSerProIle---GluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsn 19  
DB 4150 GTTGATCAAGCGTCGACAACTCAAAATCAAGCAATTGATTAATACAAACGCTCTACACT 4209

OY 20 LeuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluArgAla----- 36  
DB 4210 GAAGAGAAAATGCGACAAAGATTACTTTAAAGCTTAAAGAAAAGCGTATCAAGAT 4269  
OY 37 -----11LeaspGluLeuLysGluAlaIleGluAsp 47  
DB 4270 ATCTTAATGCAACAACAACTAATGATGTTCGCAAAATTAAGATCAAGAGTTGCTGAT 4329  
OY 48 LysGlu-----AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAla 64  
DB 4330 ATTCAGGATTAATCTACTGCAATCAACAACTTAAGATGTTCGCAAGATTAATCAACA 4389  
OY 65 LeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValAlaValLysAlaAspAsn 84  
DB 4390 AAGCAAAAGCAACAAAAGCGCTTATTCACAACTGCA-----GAT 4431  
OY 85 AlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 102  
DB 4432 GCGACTACTGAAGAAAAGAACCAATCAACAACTAGACGCAATTAACACACAGT 4491  
OY 103 -----GluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrp 118  
DB 4492 AATCAAAATATTTGAAAATGACACAGTCATGATGATTAACACTGCAAAAGATTAATCA 4551  
OY 119 GluLysAlaAlaThrProIle-----AlaLeuAspValLys----- 130  
DB 4552 ATTCAGCAATTTGACCAATTCAGATCAACAGATGTTAAACGATGCAAGACGCA 4611  
OY 130 ----- 130  
DB 4612 TTGCTAAGCAAAATGCAAAATTAATTAAGTGAATACTTAATTAATGACACTACTAAT 4671  
OY 131 -----LysThrLysAspThrLysProVal-----ValLysLysGluGluArgGlnAsn 146  
DB 4672 GAAGAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4731  
OY 147 ValAsnThrLeuProThrThrGlyGlu 155  
DB 4732 ATTAATGCAAGCAACTACTACAGTGAT 4758

RESULT 37  
ID AAS5232 standard; DNA; 7437 BP.  
XX  
XX AAS5232;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
DE Staphylococcus aureus DNA for cellular proliferation protein #1544.  
XX  
XX Antisense; ds; prokaryotic cellular proliferation gene;  
XX antibiotic; antibacterial; drug design.  
XX  
OS Staphylococcus aureus.  
XX  
PN WO200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US09180.  
XX  
XX 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;  
 XX WPI: 2001-611495/70.  
 DR P-PSDB: AAU37373.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 XX Claim 27; Seq ID No 8869; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp://wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 7437 BP; 3199 A; 1326 C; 1298 G; 1614 T; 0 other:  
 XX  
 Alignment Scores:  
 Pred. No.: 0.439 Length: 7437  
 Score: 104.00 Matches: 51  
 Percent Similarity: 37.80% Conservative: 28  
 Best Local Similarity: 24.40% Mismatches: 70  
 Query Match: 13.38% Indels: 60  
 DB: 23 Gaps: 8  
 US-09-847-539a-6 (1-159) x AAS5232 (1-7437)  
 QY 1 ValAspSerProIle---GIuGInProArgIleIleProAsnGlyThrLeuThrAsn 19  
 DB 4150 GTTGATGACAACTGTGACAACTCAAAATCAAGCAATTGATATCAACCTGCTGCTACAACT 4209  
 QY 20 LeuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluArgAla----- 36  
 DB 4210 CAAGACAAAATATCCACCAAAAGATTAGCTTTAAAGCTTAAAGCAAAAGCGTTCAAGAT 4269  
 QY 37 -----IleAspGluLeuLysLysGlnAlaIleGluAsp 47  
 DB 4270 ATCTAAATGCACAACAACATAATGATGTTACCAAAATTAAAGATCAACAGCGTTGCGAT 4329  
 QY 48 LysGlu-----AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAla 64  
 DB 4330 ATTCAAGGTATTACTGTCAGATACAAATTAAGATGTCGCAAGATGATTAATGCAACA 4389  
 QY 65 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsn 84  
 DB 4390 AAAGCAAAACGACAAAAGCGCTTATTGCAAACTGCA-----GAT 4431  
 QY 85 AlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 102  
 DB 4432 GCGACTACAGCAAGAAAAGAAAAGCAAAATCAACAGTAGACCAATTAACAAAGT 4491  
 QY 103 -----GluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTyr 118  
 DB 4492 AATCAAAATATTGAAATATGCAACAGTCATGATGTTAAACACTGCAAAAGATTAATGCA 4551  
 QY 119 GluLysAlaAlaThrProIle-----AlaLeuAspValLys----- 130  
 DB 4552 ATTCAAGCAATTGACCAACATCAAGCATCAAGATGTTAAAGCAATGCAAGACGCGAA 4611

QY 130 ----- 130  
 DB 4612 TTGCTACTGAATGCAAAATATAATAGTAACTACTAATATATGAGACTACTAAT 4671  
 QY 131 -----LysThrLysAspThrLysProVal-----ValLysLysGluGluArgGlnAsn 146  
 DB 4672 CAAGCAAAAGCTAAGTAAATTTGACACAGTTAGACCAACCATATCAACAGCTTTAAATAT 4731  
 QY 147 ValAspThrLeuProThrThrGlyGlu 155  
 DB 4732 ATTATGCAACCACTACTACAGGTGAT 4758  
 RESULT 38  
 AAS52178  
 ID AAS52178 standard; DNA: 6228 BP.  
 XX  
 AC AAS52178:  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Staphylococcus aureus DNA for cellular proliferation protein #595.  
 XX  
 KW Antisense; ds; prokaryotic cellular proliferation gene;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX 21-MAR-2001; 2001MO-US09180.  
 PF  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI: 2001-611495/70.  
 DR P-PSDB: AAU34319.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 XX Claim 27; Seq ID No 4760; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at



```

Db 3811 GATGTAACAAACAGCAGCGCGA---GATTAATCATCTGCTGAGTGGCGAAGCGTATT 3867
Oy 78 ---AlaValAlaValAlaSerAspAlaSerAspAlaLeuAlaLeuAlaAspGln 96
Db 3868 GAAGCGCTTAACAAACACCTAATGCACTGACGACAAAGACGCTGCTGTTAATCAA 3927
Oy 97 ThrAspAlaLeuGlnSerGlu-----GluAlaGln 106
Db 3928 ATCAATCACTTAAGATCAACCAATTAATCAAAATTAATCAAAACCAACCAATGATCAG 3987
Oy 107 ValValGlnSerAspAlaAlaSerAspAlaTrpGlnValAlaAlaTrpProIleAla 126
Db 3988 GTGACACACAACTCAATCAAGCGTAAATGCTATAGATTAAGTTGAAGCTGAAGTAGTA 4047
Oy 127 LeuAspValAlaValSerThrLysAspThrLysProValValLysLysGlnGluArgGlnAsn 146
Db 4048 ATTAAACCAAGGCAATTCGACGATATTGAAAAAGCTGTTAAAGAAAAAGCAGCAGCAAAATT 4107
Oy 147 ValAsnThrLeuProThrThrGlyGlnGlu 156
Db 4108 GATTAATAGTCTTGCATTCACACGATATATGAG 4137

RESULT 40
AAQ29471
ID AAQ29471 standard; DNA: 6744 BP.
XX
AC AAQ29471;
XX
DT 04-MAR-1993 (first entry)
XX
DE Extracellular factor related protein gene.
XX
KW EF*; detection; prevention; screening; diagnostic; ss.
XX
OS Streptococcus suis type II (non-pathogenic).

Location/Qualifiers
FH CDS 361..5826
FT /tag= a
FT sig_peptide 361..498
FT /tag= b
FT mat_peptide 499..5826
FT /tag= c
FT promoter 66..71
FT /tag= d
FT promoter 89..94
FT /tag= e
FT promoter 153..158
FT /tag= f
FT promoter 176..181
FT /tag= g
FT RBS 350..356
FT /note= "-10 region"
FT /tag= h
FT /note= "ribosome binding site"
FT /tag= i
FT repeat_unit 2869
FT /note= "start of repetitive unit R1"
FT /tag= j
FT repeat_unit 3097
FT /note= "start of repetitive unit R2"
FT /tag= k
FT repeat_unit 3292
FT /note= "start of repetitive unit R3"
FT /tag= l
FT repeat_unit 3520
FT /note= "start of repetitive unit R4"
FT /tag= m
FT repeat_unit 4087
FT /note= "start of repetitive unit R5"

```

```

FT repeat_unit 4381
FT /tag= n
FT /note= "start of repetitive unit R6"
FT repeat_unit 4609
FT /tag= o
FT /note= "start of repetitive unit R7"
FT repeat_unit 4837
FT /tag= p
FT /note= "start of repetitive unit R8"
FT repeat_unit 5065
FT /tag= q
FT /note= "start of repetitive unit R9"
FT repeat_unit 5293
FT /tag= r
FT /note= "start of repetitive unit R10"
FT repeat_unit 5521
FT /tag= s
FT /note= "start of repetitive unit R11"
FT misc_feature 6554..6566
FT /tag= t
FT /note= "dyad symmetry region 1a"
FT misc_feature 6571..6583
FT /tag= u
FT /note= "dyad symmetry region 1b"
FT misc_feature 6611..6625
FT /tag= v
FT /note= "dyad symmetry region 2a"
FT misc_feature 6631..6644
FT /tag= w
FT /note= "dyad symmetry region 2b"

```

```

WO9216630-A.
XX
PD 01-OCT-1992.
XX
PF 19-MAR-1992; 92MO-NL00054.
XX
PR 21-MAR-1991; 91NL-0000510.
XX
PA (DIER-) GENT DIERGENESKUNDIG INST.
XX
PI Smitch HE, Vecht U;
XX
DR WPI: 1992-349215/42.
XX
P-PSDB; AAR27745.
XX
PT Deoxyribonucleic acid encoding virulence characteristic of
XX Streptococcus suis - useful for antibody and polypeptide for
XX diagnosing and preventing infections in pigs and humans
XX
PS Claim 2; Fig 1b; 86pp; English.
XX
CC The sequence is that of the extracellular factor related protein gene
CC from Streptococcus suis type II (non-pathogenic) which allows the
CC detection and the prevention of infections by S. suis in a more
CC effective manner than was previously possible. It facilitates screening
CC of e.g. pigs and elimination of infected and carrier pigs can then be
CC carried out. The new diagnostic tests can distinguish between
CC avirulent and virulent strains. It may be used in the prodn. of
CC a vaccine. See also AAQ29470 and AAQ29472.
XX
SQ Sequence 6744 BP; 2402 A; 1231 C; 1535 G; 1576 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 0.441
Score: 103.50
Percent Similarity: 46.85%
Best Local Similarity: 24.48%
Query Match: 13.32%
DB: 13
Gaps: 5

```

US-09-847-539a-6 (1-159) x AAQ29471 (1-6744)



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:04:37 : Search time 64.6203 Seconds  
(Without alignments) 273.300 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQPRIPNGTILNL.....KKEERQVNTLPPTGESNP 159

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

1:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.*
2:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.*
3:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.*
4:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.*
5:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.*
6:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.*
7:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.*
8:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.*
9:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.*
10:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.*
11:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.*
12:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.*
13:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.*
14:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.*
15:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.*
16:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.*
17:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.*
18:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.*
19:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.*
20:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.*
21:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.*
22:	/SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	777	100.0	159	21	AAV71041
2	777	100.0	184	21	AAV71039
3	777	100.0	217	21	AAV71042
4	760	97.8	167	21	AAV71046
5	750	96.5	155	21	AAV71044
6	694	89.3	271	21	AAV71045
7	682	89.1	259	21	AAV71043
8	682	87.8	141	21	AAV71040
9	285	36.7	58	21	AAV71037
10	188	24.2	448	8	AAV70493
11	188	24.2	448	10	AAV95030

12	188	24.2	448	11	AAV7013	Protein G variant
13	188	24.2	448	12	AAV10004	Streptococcus GX78
14	188	24.2	448	15	AAV33290	Streptococcus Prot
15	187	24.1	593	11	AAV7014	Protein G variant
16	187	24.1	593	15	AAV62944	Streptococcus Prot
17	187	24.1	594	12	AAV10005	Streptococcus GX78
18	142	18.3	103	21	AAV57611	Streptococcus stra
19	139	17.9	235	10	AAV94785	Protein G variant.
20	139	17.9	235	11	AAV7004	Protein G variant.
21	139	17.9	265	12	AAV10011	Type 4 GX7809 prot
22	139	17.9	265	15	AAV33294	IgG-binding Strept
23	129	16.6	28	21	AAV71038	Streptococcus pyog
24	120	15.4	23	21	AAV71036	Streptococcus pyog
25	116	14.9	166	21	AAV38475	Arbidiopsis thalia
26	115	14.8	25	21	AAV71048	Streptococcus pyog
27	114.5	14.7	413	16	AAV71928	S. dysgalactiae MA
28	114.5	14.7	413	16	AAV71670	S. dysgalactiae MA
29	112	14.4	154	21	AAV38476	Arbidiopsis thalia
30	112	14.4	480	8	AAV70468	Sequence of polype
31	109	14.0	191	21	AAV41392	Arbidiopsis thalia
32	109	14.0	230	21	AAV41381	Arbidiopsis thalia
33	109	14.0	316	21	AAV41390	Arbidiopsis thalia
34	108.5	14.0	180	21	AAV20162	Arbidiopsis thalia
35	108	13.9	1096	22	AAV62318	Drosophila melanog
36	105.5	13.6	168	21	AAV20163	Arbidiopsis thalia
37	105.5	13.6	314	18	AAV27286	Rabbit enteropatho
38	105	13.5	208	16	AAV71127	Synthetic protein
39	105	13.5	664	16	AAV71929	S. dysgalactiae MI
40	104.5	13.4	151	21	AAV37153	Arbidiopsis thalia
41	104	13.4	2478	22	AAV34320	Staphylococcus aur
42	104	13.4	2478	22	AAV37374	Staphylococcus aur
43	103.5	13.3	1822	13	AAV27745	Extracellular fact
44	103.5	13.3	2076	22	AAV34319	Staphylococcus aur
45	103.5	13.3	2186	22	AAV37320	Staphylococcus aur

## ALIGNMENTS

RESULT 1	AAV71041	standard; peptide: 159 AA.
ID	AAV71041	
XX	AAV71041	
AC	AAV71041	
XX		
DT	29-AUG-2000	(first entry)
XX		
XX		
DE	Streptococcus pyogenes strain SF370 GRAB protein fragment #5.	
XX		
XX		
KW	GRAB protein; protein G related alpha2m binding protein; vaccine;	
KW	alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;	
KW	Immune response; Streptococcus pyogenes infection.	
OS	Streptococcus pyogenes.	
XX		
PN	W0200026240-A2.	
XX		
PD	11-MAY-2000.	
XX		
PF	02-NOV-1999;	99WO-GB03631.
XX		
PR	02-NOV-1998;	98GB-0023975.
XX		
PA	(ACT-) ACTINOVA LTD.	
XX		
PI	Bjorck LH, Rasmussen M;	
XX		
DR	WPI; 2000-365572/31.	
XX		
PT	New alpha2m binding protein for generating a protective immune response	
PT	to group A streptococcus and purifying the binding protein	
XX		
PS	Claim 5; Page 57-58; 67pp; English.	

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-192 and devoid of the membrane spanning region. This  
CC fragment is useful in vaccine composition.  
XX  
SQ Sequence 159 AA:  
  
Query Match 100.0%; Score 777; DB 21; Length 159;  
Best Local Similarity 100.0%; Pred. No. 1.2e-61;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDSPIDPRLIPNGGTLTNLGNAPKELALRNEERAIDELKKAIEDKETTATIEAASSD 60  
Db 1 VDSPIDPRLIPNGGTLTNLGNAPKELALRNEERAIDELKKAIEDKETTATIEAASSD 60  
  
QY 61 ALFALADOTDALOSEEAAYVKAADNAASDLEALADOTDALOSEEAAYVQSDNNAASDAWEX 120  
Db 61 ALFALADOTDALOSEEAAYVKAADNAASDLEALADOTDALOSEEAAYVQSDNNAASDAWEX 120  
  
QY 121 AATPIALDVKKTKDTKRPVVKKEERQNVNTLPTTGEESNP 159  
Db 121 AATPIALDVKKTKDTKRPVVKKEERQNVNTLPTTGEESNP 159  
  
RESULT 2  
AA71039  
ID AAY71039 standard; Protein; 184 AA.  
XX  
AC AAY71039;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 mature GRAB protein.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW Immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
FH Key Location/Qualifiers  
FT Binding-site 1..58 /label= alpha2-macroglobulin\_binding-site  
FT Region 59..86 /label= R1  
FT /note= "repeat region"  
FT 87..114 /label= R2  
FT /note= "repeat region"  
FT 115..159 /label= Cell\_wall\_spanning\_region  
FT Region 150..155 /label= Cell\_wall\_spanning\_region  
FT /note= "consensus sequence for gram-positive  
FT surface cell wall anchored proteins"  
FT 160..184 /label= Membrane\_spanning\_region  
FT Region  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PP 02-NOV-1999; 99WO-CB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX

PA (ACT1-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Raasmussen M;  
PI WPL; 2000-365572/31.  
XX  
DR  
XX  
PT- New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein  
PS Claim 5; Page 56; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a mature GRAB protein from S. pyogenes strain SF370 without the  
CC signal sequence. This sequence is capable of binding alpha2M  
CC and useful in vaccine composition.  
XX  
SQ Sequence 184 AA:  
  
Query Match 100.0%; Score 777; DB 21; Length 184;  
Best Local Similarity 100.0%; Pred. No. 1.4e-61;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDSPIDPRLIPNGGTLTNLGNAPKELALRNEERAIDELKKAIEDKETTATIEAASSD 60  
Db 1 VDSPIDPRLIPNGGTLTNLGNAPKELALRNEERAIDELKKAIEDKETTATIEAASSD 60  
  
QY 61 ALFALADOTDALOSEEAAYVKAADNAASDLEALADOTDALOSEEAAYVQSDNNAASDAWEX 120  
Db 61 ALFALADOTDALOSEEAAYVKAADNAASDLEALADOTDALOSEEAAYVQSDNNAASDAWEX 120  
  
QY 121 AATPIALDVKKTKDTKRPVVKKEERQNVNTLPTTGEESNP 159  
Db 121 AATPIALDVKKTKDTKRPVVKKEERQNVNTLPTTGEESNP 159  
  
RESULT 3  
AA71042  
ID AAY71042 standard; Protein; 217 AA.  
XX  
AC AAY71042;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 full-length GRAB protein.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW Immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..33 /label= Signal\_sequence  
FT Protein 34..217 /label= Mature\_GRAB\_protein  
FT Binding-site 34..91 /label= alpha2M\_binding-site  
FT 34..58 /label= alpha2M\_binding-site  
FT Domain 92..119 /note= "shows homology to E domain of protein G"  
FT Region 120..147 /label= Repeat\_region\_1  
FT Region 148..192 /label= Repeat\_region\_2  
FT /label= Cell\_wall\_spanning\_region  
FT

FT	Region	183..188
FT	/note= "consensus sequence for gram-positive surface cell wall anchored proteins"	
FT	Region	193..217
FT	/label= Membrane-spanning_region	
PX	MO200026240-A2.	
PD	11-MAY-2000.	
XX		
XX	02-NOV-1999;	99WO-CB03631.
XX		
XX	02-NOV-1998;	98GB-0023975.
XX		
PA	(ACTI-) ACTINOVA LTD.	
XX		
PI	Bjorck LH, Rasmussen M;	
XX		
XX	WPI: 2000-365572/31.	
DR	N-PSDB; AAD00559; AAD00560.	
XX		
PT	New alpha2M binding protein for generating a protective immune response to group A streptococcus and purifying the binding protein	
XX		
PS	Claim 5; Fig 2B; 67pp; English.	
CC	The patent discloses a new family of proteins termed GRAB (protein G related alpha2M binding protein) from Streptococcus pyogenes which have the ability to bind alpha2-macroglobulin (alpha2M) and show homology to protein G of group G Streptococcus. GRAB protein and peptides derived from it are used in vaccine compositions for generating a protective immune response against group A Streptococcus. Antibodies against GRAB are useful for treating Streptococcus pyogenes infections. The protein is also useful for purifying alpha2M from a sample. The present sequence is a full-length GRAB protein from S. pyogenes strain SF370.	
SQ	Sequence	217 AA:
Query Match	100.0%;	Score 777; DR 21; Length 217;
Best Local Similarity	100.0%;	Pred. No. 1,7e-61;
Matches 159; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
OY	1 VDSPEOPRIIPNGGTITNLGNAPKIALRNEERAIIDELKKQATIEDKATATAIEAASD	60
DB		
DB	34 VDSPLEQPPIIPNGGTTNLTGNAPKIALRNEERAIDELKKQATIEDKATATAIEAASD	93
OY	61 ALFALADDTDALQSEEAFAVKADNNAASDALFALADDTDALQSEEAFAVQSDNNAASDAWEK	120
DB	94 ALFALADDTDALQSEEAFAVKADNNAASDLAEFLADDTDALQSEEAFAVQSDNNAASDAWEK	153
OY	121 AATPIALDVKKTKDKTPPVKKKEEKNVNTLPTTGESNP	159
DB		
DB	154 AATPIALDVKKTKDKTPPVKKKEEKNVNTLPTTGESNP	192
RESULT 4		
ID	AA71046	
ID	AA71046 standard; Protein; 167 AA.	
AC	AA71046;	
XX		
DT	29-AUG-2000 (first entry)	
XX		
DE	Streptococcus pyogenes strain KTL3 partial GRAB protein.	
XX		
KW	GRAB protein; protein G related alpha2M binding protein; vaccine; alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic; immune response; Streptococcus pyogenes infection.	
XX		
OS	Streptococcus pyogenes.	
PX		
PX	WO200026240-A2.	
XX		

PD	11-MAY-2000.
XX	
PF	02-NOV-1999; 99WO-GB03631.
XX	
PR	02-NOV-1998; 98GB-0023975.
XX	
PA	(ACTI-) ACTINOVA LTD.
XX	
PI	Bjorck LH, Rasmussen M;
DR	WPI: 2000-365572/31.
N-PSDB:	AAD00564.
XX	
PT	New alpha2m binding protein for generating a protective immune response
PT	to group A streptococcus and purifying the binding protein -
XX	
PS	Claim 5; Page 62-63; 67pp; English.
XX	
CC	The patent discloses a new family of proteins termed GRAB (protein G
CC	related alpha2m binding protein) from Streptococcus pyogenes which have
CC	the ability to bind alpha2-macroglobulin (alpha2m) and show homology to
CC	protein G of group G Streptococcus. GRAB protein and peptides derived
CC	from it are used in vaccine compositions for generating a protective
CC	immune response against group A Streptococcus. Antibodies against GRAB
CC	are useful for treating Streptococcus pyogenes infections. The protein
CC	is also useful for purifying alpha2m from a sample. The present sequence
CC	is a partial GRAB protein from S. pyogenes strain KTL3.
CC	The protein has alpha2m binding region and is useful in vaccine
CC	composition.
XX	
SQ	Sequence 167 AA:
Query Match	97.8%; Score 760; DB 21; Length 167;
Best Local Similarity	100.0%; Pred. No. 4e-60;
Matches 156; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 VDSPEOPRIIPNGSTIRNLGNAPEKIALRNEFAIDELKCAIEDKATTAIEAASD 60
DB	12 VDSPLEQPIIPINGGLTNLGNAPKIALRNEERAIDELKCAIEDKATTATIEAASSD 71
OY	61 ALEALADDTALQSEEAFAAVVKADNNAASDALAEALADDTALQSEEAFAVVSDNAASDAWEK 120
DB	72 ALEALADDTALQSEEAFAAVVKADNNAASDALAEALADDTALQSEEAFAVVSDNAASDAWEK 131
OY	121 AATPIALDVKKTKDKTPPVYKKKEERONVNTLPPTGEE 156
DB	132 AATPIALDVKKTKDKTPPVYKKKEERONVNTLPPTGEE 167
RESULT 5	
AA771044	
ID	AA771044 standard; Protein; 155 AA.
XX	
AC	AA771044;
XX	
DT	29-AUG-2000 (first entry)
XX	
DE	Streptococcus pyogenes strain API partial GRAB protein.
XX	
KW	GRAB protein; protein G related alpha2m binding protein; vaccine;
RW	alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW	immune response; Streptococcus pyogenes infection.
XX	
OS	Streptococcus pyogenes.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 17
FT	/note-"Thr at position 18 of GRAB protein fragment
FT	(AA771036) from S. pyogenes strain SF370 is replaced
FT	with Ile"
XX	
PN	MO200026240-A2.
XX	

PD 11-MAY-2000.  
XX  
XX 02-NOV-1999; 99WO-GB03631.  
XX  
XX 02-NOV-1998; 98GB-0023975.  
XX  
XX (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
XX  
XX WPI: 2000-365572/31.  
XX  
XX N-PSDB: AAD00562.  
XX  
XX New alpha2m binding protein for generating a protective immune response  
XX  
XX to group A streptococcus and purifying the binding protein -  
XX  
XX Claim 5; Page 60-61; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
XX  
XX related alpha2m binding protein) from Streptococcus pyogenes which have  
XX  
XX the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
XX  
XX protein G of group G Streptococcus. GRAB protein and peptides derived  
XX  
XX from it are used in vaccine compositions for generating a protective  
XX  
XX immune response against group A Streptococcus. Antibodies against GRAB  
XX  
XX are useful for treating Streptococcus pyogenes infections. The protein  
XX  
XX is also useful for purifying alpha2m from a sample. The present sequence  
XX  
XX is a partial GRAB protein from S. pyogenes strain AP1.  
XX  
XX The protein has alpha2m binding region and is useful in vaccine  
XX  
XX composition.  
XX  
XX  
SQ Sequence 155 AA:  
Query Match 96.5%; Score 750; DB 21; Length 155;  
Best Local Similarity 99.4%; Pred. No. 2.9e-59;  
Matches 154; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 DSPTEOPRIIPNGGTLTNLGNAPKALRNERRAIDELKKAIDKEATTATIEAASSDA 61  
DB 1 DSPTEOPRIIPNGGTLTNLGNAPKALRNERRAIDELKKAIDKEATTATIEAASSDA 60  
QY 62 LPAIDQDTALQSEEAAYVKADNAASDALEALADQDTALQSEEAAYVSDNNAASDAWEKA 121  
DB 61 LEALADQDTALQSEEAAYVKADNAASDALEALADQDTALQSEEAAYVSDNNAASDAWEKA 120  
QY 122 ATPIALDVAKTKDTPVVKKEERQNVNTLPTTGE 156  
DB 121 ATPIALDVAKTKDTPVVKKEERQNVNTLPTTGE 155  
RESULT 6  
AAV71045  
ID AAV71045 standard; Protein; 271 AA.  
XX  
XX AAV71045;  
XX  
XX 29-AUG-2000 (first entry)  
XX  
XX Streptococcus pyogenes strain AP49 partial GRAB protein.  
XX  
XX  
XX GRAB protein; protein G related alpha2m binding protein; vaccine;  
XX  
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
XX  
XX Immune response; Streptococcus pyogenes infection.  
XX  
XX Streptococcus pyogenes.  
XX  
XX OS  
XX PN WO200026240-A2.  
XX  
XX 11-MAY-2000.  
XX  
XX PD 02-NOV-1999; 99WO-GB03631.  
XX  
XX PF 02-NOV-1998; 98GB-0023975.  
XX  
XX PR  
XX

PA (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
XX  
XX WPI: 2000-365572/31.  
XX  
XX N-PSDB: AAD00563.  
XX  
XX New alpha2m binding protein for generating a protective immune response  
XX  
XX to group A streptococcus and purifying the binding protein -  
XX  
XX Claim 5; Page 61-62; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
XX  
XX related alpha2m binding protein) from Streptococcus pyogenes which have  
XX  
XX the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
XX  
XX protein G of group G Streptococcus. GRAB protein and peptides derived  
XX  
XX from it are used in vaccine compositions for generating a protective  
XX  
XX immune response against group A Streptococcus. Antibodies against GRAB  
XX  
XX are useful for treating Streptococcus pyogenes infections. The protein  
XX  
XX is also useful for purifying alpha2m from a sample. The present sequence  
XX  
XX is a partial GRAB protein from S. pyogenes strain AP49.  
XX  
XX The protein has alpha2m binding region and is useful in vaccine  
XX  
XX composition.  
XX  
XX  
SQ Sequence 271 AA:  
Query Match 89.3%; Score 694; DB 21; Length 271;  
Best Local Similarity 58.2%; Pred. No. 5.5e-54;  
Matches 156; Conservative 0; Mismatches 0; Indels 112; Gaps 1;  
QY 1 VDSPTEOPRIIPNGGTLTNLGNAPKALRNERRAIDELKKAIDKEATTATIEAAS-- 58  
DB 4 VDSPTEOPRIIPNGGTLTNLGNAPKALRNERRAIDELKKAIDKEATTATIEAASD 63  
QY 59 -----  
DB 64 ALBALADQDALQSEEAAYVSDNNAASDALEALADQDALQSEEAAYVSDNNAAGDALEA 123  
QY 59 -----SDALEALADQ 68  
DB 124 LADQDTALQSEEAAYVKADNAASDALEALADQDTALQSEEAAYVKADNAASDALEALADQ 183  
QY 69 TDALQSEEAAYVKADNAASDALEALADQDTALQSEEAAYVSDNNAASDAWEKAATP1AID 128  
DB 184 TDALQSEEAAYVKADNAASDALEALADQDTALQSEEAAYVSDNNAASDAWEKAATP1AID 243  
QY 129 VKTKDTPVVKKEERQNVNTLPTTGE 156  
DB 244 VKTKDTPVVKKEERQNVNTLPTTGE 271  
RESULT 7  
AAV71043  
ID AAV71043 standard; Protein; 259 AA.  
XX  
XX AAV71043;  
XX  
XX 29-AUG-2000 (first entry)  
XX  
XX Streptococcus pyogenes strain KTL9 partial GRAB protein.  
XX  
XX  
XX GRAB protein; protein G related alpha2m binding protein; vaccine;  
XX  
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
XX  
XX Immune response; Streptococcus pyogenes infection.  
XX  
XX Streptococcus pyogenes.  
XX  
XX OS  
XX PN WO200026240-A2.  
XX  
XX 11-MAY-2000.  
XX  
XX PD 02-NOV-1999; 99WO-GB03631.  
XX  
XX PF  
XX

PR 02-NOV-1998; 98GB-0023975.  
XX  
XX (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
PI WPI; 2000-365572/31.  
DR N-PSDB; AAD00561.  
XX  
PT New alpha2m binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
XX  
PS Claim 5; Page 59-60; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2m binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2m from a sample. The present sequence  
CC is a partial GRAB protein from S. pyogenes strain K7L9.  
CC The protein has alpha2m binding region and is useful in vaccine  
CC composition.  
XX  
XX Sequence 259 AA:  
SQ  
Query Match 89.1%; Score 692; DB 21; Length 259;  
Best Local Similarity 64.0%; Pred. No. 7,8e-54;  
Matches 153; Conservative 0; Mismatches 2; Indels 84; Gaps 1;  
QY 1 VDSPIEOPRIIPNGGTLNLGNAPKALRNERAIDELKKAIDEKATTAIEAASD 60  
DB 21 VDSPIEOPRIIPNGGTLNLGNAPKALRNERAIDELKKAIDEKATTAIEAASD 80  
QY 61 A----- 61  
DB 81 ALEALADQADALQSEEAAYVQSDNNAADALEALADQTDALQSEEAAYVKADNNAADTLEA 140  
QY 62 -----LEALADQTDALQSEEAAYVKADNNAADALEALADQ 96  
DB 141 LADQTDALQSEEAAYVKADNNAADTLEALADQTDALQSEEAAYVKADNNAADTLEALADQ 200  
QY 97 TDAIQSEEAAYVQSDNNAADNKAATPIALDVKKTKDTKPPVKKERQVNTLPTTGE 155  
DB 201 TDAIQSEEAAYVQSDNNAADNKAATPIALDVKKTKDTKPPVKKERQVNTLPTTGE 259  
RESULT 8  
AA71040  
ID AA71040 standard; peptide; 141 AA.  
XX  
XX AA71040;  
AC  
XX  
XX 29-AUG-2000 (first entry)  
DT  
XX  
XX Streptococcus pyogenes strain SF370 GRAB protein fragment #4.  
DE  
XX  
XX GRAB protein; protein G related alpha2m binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antidiabetic;  
KM Immune response; Streptococcus pyogenes infection.  
XX  
XX Streptococcus pyogenes.  
OS  
XX  
XX WO200026240-A2.  
PN  
XX  
XX 11-MAY-2000.  
PD  
XX  
XX 02-NOV-1999; 99MO-GB03631.  
PF  
XX  
XX 02-NOV-1998; 98GB-0023975.  
PR  
XX  
XX

PA (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
PI WPI; 2000-365572/31.  
DR  
XX  
XX New alpha2m binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
XX  
PS Claim 5; Page 57; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2m binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2m from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-174. This fragment is devoid of the trans-membrane  
CC and cell wall anchor regions. It is useful in vaccine composition.  
XX  
XX Sequence 141 AA:  
SQ  
Query Match 87.8%; Score 682; DB 21; Length 141;  
Best Local Similarity 100.0%; Pred. No. 2,8e-53;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDSPIEOPRIIPNGGTLNLGNAPKALRNERAIDELKKAIDEKATTAIEAASD 60  
DB 1 VDSPIEOPRIIPNGGTLNLGNAPKALRNERAIDELKKAIDEKATTAIEAASD 60  
QY 61 ALEALADQADALQSEEAAYVKADNNAADALEALADQTDALQSEEAAYVQSDNNAADNKA 120  
DB 61 ALEALADQADALQSEEAAYVKADNNAADALEALADQTDALQSEEAAYVQSDNNAADNKA 120  
QY 121 AAPPIALDVKKTKDTKPPVKK 141  
DB 121 AAPPIALDVKKTKDTKPPVKK 141  
RESULT 9  
AA71037  
ID AA71037 standard; peptide; 58 AA.  
XX  
XX AA71037;  
AC  
XX  
XX 29-AUG-2000 (first entry)  
DT  
XX  
XX Streptococcus pyogenes strain SF370 GRAB protein fragment #2.  
DE  
XX  
XX GRAB protein; protein G related alpha2m binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antidiabetic;  
KM Immune response; Streptococcus pyogenes infection.  
XX  
XX Streptococcus pyogenes.  
OS  
XX  
XX WO200026240-A2.  
PN  
XX  
XX 11-MAY-2000.  
PD  
XX  
XX 02-NOV-1999; 99MO-GB03631.  
PF  
XX  
XX 02-NOV-1998; 98GB-0023975.  
PR  
XX  
XX (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
PI  
XX  
XX WPI; 2000-365572/31.  
DR  
XX  
XX New alpha2m binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 2: Page 55: 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2m binding protein) from streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
CC protein G of group G streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2m from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-91. This fragment is capable of binding alpha2m  
XX and useful in vaccine composition.  
SO Sequence 58 AA:  
Query Match 36.7%; Score 285; DB 21; Length 58;  
Best Local Similarity 100.0%; Pred. No. 2e-18;  
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDSPIEQPIIRNGGTLNLGNAPKALRNERRAIDELKKQALDEKATTAIEAAS 58  
DB 1 VDSPIEQPIIRNGGTLNLGNAPKALRNERRAIDELKKQALDEKATTAIEAAS 58  
RESULT 10  
AAP70493  
ID AAP70493 standard: protein; 448 AA.  
XX  
AC AAP70493;  
XX  
DT 06-MAR-1991 (first entry)  
XX  
DE Protein G.  
XX  
KM Protein G; antibody; Fc receptor;  
XX  
OS Streptococcus Lancefield Group G strain.  
XX  
FH Key Location/Qualifiers  
FT Active-site 228..297  
FT /label=active site B1 298..352  
FT Active-site /label=active site B2  
XX  
PN WO8705025-A.  
XX  
PD 27-AUG-1987.  
XX  
PF 17-FEB-1987; 87WO-US00329.  
XX  
PR 22-APR-1986; 86US-0854997.  
PR 14-FEB-1986; 86US-0829354.  
XX  
PA (GENE-) GENEX CORP.  
PA (FAHN-) FAHNSTOCK S R.  
XX  
PI Fahnestock S;  
XX  
DR WPI; 1987-250197/35.  
DR N-PSDB; AAN70811.  
XX  
PT Cloned Protein G gene - used for producing Protein G for  
XX detection and purification. of antibodies and treatment of diseases  
PS Disclosure; Fig. 3; 68pp; English.  
XX  
CC Protein G expressed by inserting the gene into an expression  
CC vector. A second vector may also be used as a cryptic helper plasmid  
CC to stably maintain the first plasmid in the host cell. Bacterial Fc  
CC receptors such as Protein G can be used to detect and purify

CC antibodies, and in the treatment of disease. Fc receptors are useful  
CC to purify antibodies to be used in the purification of protein drugs and  
CC as therapeutics. High levels of Protein G can be obd. in conditions  
CC favourable for isolation, using a non-pathogenic host. Suitable  
CC cloning vectors are lambda gt11, M13mp9 and pGX1066.  
XX  
SO Sequence 448 AA:  
Query Match 24.2%; Score 188; DB 8; Length 448;  
Best Local Similarity 38.9%; Pred. No. 1e-08;  
Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;  
QY 1 VDSPIEQPIIRNGGTLNLGNAPKALRNERRAIDELKKQALDEKATTAIEAASD 60  
DB 34 VDSPIEDPTIIRNGGELTNLGNSETTLALRNESATADLTAAAVDTVAAMAAEYACAA 93  
QY 61 ALEALADPTDAQSEEAAYKADN--AASDALEAL--ADQTDALQSEAEVQSSNMA- 114  
DB 94 AWEAALA-AADALAKAKADLKEFNKYGVDYKLNLNKKTVEGJLDLQAOVESAKKAR 152  
QY 115 -SDANE-----KAATPIALDVKKTK--DTKPVVKE 142  
DB 153 ISEATDGLSDFLKSGTPAEDTVKSTIELAKYLANRE 189  
RESULT 11  
AAP95030  
ID AAP95030 standard: protein; 448 AA.  
XX  
AC AAP95030;  
XX  
DT 04-JUL-1990 (first entry)  
XX  
DE Protein G.  
XX  
KM Protein G; immunoglobulin; Fc receptor; ds.  
XX  
OS Streptococcus sp.  
XX  
PN WO8810306-A.  
XX  
PD 29-DEC-1988.  
XX  
PF 20-JUN-1988; 88WO-US02084.  
XX  
PR 19-JUN-1987; 87US-0063959.  
XX  
PA (GENE-) GENEX CORP.  
XX  
PI Fahnestock SR;  
XX  
DR WPI; 1989-023848/03.  
DR N-PSDB; AAN91093.  
XX  
PT Cloned protein G variant genes -  
XX expressing proteins having immunoglobulin-binding properties of  
XX protein G and derived from Streptococcus sp.  
XX  
PS Disclosure; 116pp; English.  
XX  
CC Protein G of non-pathogenic streptococcus and variants may be isolated,  
CC useful as bacterial Fc receptors eg in purification and detection of Abs.  
CC screening of hybridoma clones and treatment of disease.  
XX  
SO Sequence 448 AA:  
Query Match 24.2%; Score 188; DB 10; Length 448;  
Best Local Similarity 38.9%; Pred. No. 1e-08;  
Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;  
QY 1 VDSPIEQPIIRNGGTLNLGNAPKALRNERRAIDELKKQALDEKATTAIEAASD 60  
DB 34 VDSPIEDPTIIRNGGELTNLGNSETTLALRNESATADLTAAAVDTVAAMAAEYACAA 93

```

OY      61 ALEALADDTDALQSEEAAYVKADN--AASDALEAL---ADOTDALQSEAEVVGSDMA- 114
       | | | | | | | | | : | | | | | : | : | : | : | | | |
Db      94 AMEAAA-AADALAKAKKAAIAKEFNKYGVSDYYKKLLINNAKTVEGIKDLQAQVYESAKAR 152
OY      115 -SDAME-----KAATPIALDVKKTK--DTRPVYKKE 142
       | : | : | | | | | : | | | | | : | : | : | : |
Db      153 ISEAFTDGLSDFLKSQTPAEDTYKSTIELAEAKVLANRE 189

RESULT 12
AAR07013 standard; protein: 448 AA.
XX      AAR07013;
AC      AAR07013;
XX      17-JAN-1991 (first entry)
XX      Protein G variant with two active sites.
XX      Immunoglobulin.
XX      Streptococcus sp. Lancefield Group G strain.
OS      XX
FH      Key Location/Qualifiers
FH      Active-site 228..282
FT      /label=B1
FT      Active-site 298..352
FT      /label=B2
XX      US4956296-A.
XX      PD
XX      11-SEP-1990.
XX      PF
XX      20-JUN-1988; 88US-0209236.
XX      PR
XX      20-JUN-1988; 88US-0209236.
XX      PR
XX      14-FEB-1986; 86US-0829354.
XX      PR
XX      23-APR-1986; 86US-0854887.
XX      PR
XX      17-FEB-1987; 87MO-US00328.
XX      PR
XX      19-JUN-1987; 87US-0063959.
XX      PA
XX      (GENE-) GENEX CORP.
XX      PI
XX      Fahnestock SR.
XX      DR
XX      WPI: 1990-297491/39.
XX      DR
XX      N-PSDB: AA006018.
XX      PT
XX      Recombinant Protein G variants - obtd. using a cloned gene
XX      encoding Protein G from Streptococcus sp., used for binding
XX      immunoglobulin.
XX      PS
XX      Disclosure: Fig 8a-c: 48pp; English.
XX      CC
XX      Fragments and variants of the sequence are claimed esp. where
XX      incorporated into a non-pathogenic host eg. E.coli, and expressed
XX      at high levels.
XX      CC
XX      The variants have a higher binding efficiency and capacity for
XX      immunoglobulin, and may be used for purifying, detecting and
XX      isolating antibodies.
XX      CC
XX      SO
XX      Sequence 448 AA;

Query Match 24.2%; Score 188; DB 11; Length 448;
Best Local Similarity 38.9%; Pred. No. 1e-06;
Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps

OY      1 VDSPEOPRIIINGGTLLNLGNAPDEKALRNREERAIDELKKQAIEDKEATTATFAASD 60
       ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db      34 VDSPEDIPITINGGELTNLGNSETTTLALRNREGSATADLTAAAVADITYAAAAAENAGAA 93
OY      61 ALEALADDTDALQSEEAAYVKADN--AASDALEAL---ADOTDALQSEAEVVGSDMA- 114
```

[illegible]

```

Db      153 ISEATDGLSDPLKSGTPADPYKSIELAEKXVLANRE 189

RESULT 14
ID      AAR53290
AC      AAR53290 standard; Protein; 448 AA.
XX
XX      AAR53290;
XX
XX      06-JAN-1995 (first entry)
XX
XX      Streptococcus Protein G derived from strain GX7809.
XX
XX      Streptococcus Protein G; variant; Igg binding activity;
XX      Immunoglobulin; Lancefield Group G; bacterial Fc receptor.
XX
XX      Streptococcus sp. GX7809.
XX
XX
XX      Key      Location/Qualifiers
XX      Active-site 228..282
XX                  /label= B1
XX      Region      283..297
XX                  /label= b
XX                  /note= "linking region"
XX      Active-site 298..352
XX                  /label= B2
XX
XX      US5312901-A.
XX
XX      17-MAY-1994.
XX
XX      14-FEB-1986; 86US-0829354.
XX
XX      14-FEB-1986; 86US-0829354.
XX
XX      14-FEB-1986; 86US-0829354.
XX
XX      23-APR-1986; 86US-0854887.
XX
XX      19-JUN-1987; 87US-0063959.
XX
XX      20-JUN-1988; 88US-0209236.
XX
XX      19-JUN-1990; 90US-0540169.
XX
XX      21-APR-1992; 92US-0871539.
XX
XX      (PHNA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX
XX      Fahnstock SR:
XX
XX      WPI: 1994-159179/19.
XX
XX      N-PSDB; AAQ64644.
XX
XX      New recombinant streptococcal protein G variants - useful for
XX      antibody detection and purification and for therapy
XX
XX      Example 2; Fig 3 and Fig 8; 48bp; English.
XX
XX      A 1.9kb HindIII fragment containing the entire coding sequence for
XX      protein G was isolated from Streptococcus GX7809. The protein G has
XX      Igg-binding activity which has been localised to the B repeating
XX      structure. Streptococcal Protein G variants comprising the B domains
XX      are claimed.
XX
XX      Sequence 448 AA;
XX
XX      Query Match 24.2%; Score 188; DB 15; Length 448;
XX      Best Local Similarity 38.9%; Pred. No. 1e-08;
XX      Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps
XX
XX      1 VDSPIEGPRITPNGSTITLGNAREKIALRNEERAIDELKKQATIEDKATTAIEASSD 60
XX      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      34 VDSPIEDPTILIRNGSELNLTGNSETTIALRNEESATADTAAVAADTVAAAAAENAGAA 93
XX
XX      61 ALEAADQTDALOSEEAAYVKKADN--AASDALEAL---ADQTDALOSEEAAYVOSDMAA- 114
XX      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      94 AWEAALA--AADALAKKADALKEFNKYGSDYTKNLINNAKTYEGTKIDIAQVYVESAKKAR 152
XX
XX      115 -SDAVE-----KAATPIALDVKKTK--DTKPVVKE 142

```

Db	153	ISEATDGLSDFTLKSOTPAEDTVKSIETLAEKVLANRE	189
RESULT 15			
ID	AAR07014	standard; protein: 593 AA.	
AC	AAR07014;		
XX			
XX			
DT	17-JAN-1991	(first entry)	
XX			
DE	Protein G variant with three active sites.		
XX			
KW	Immunoglobulin.		
XX			
OS	Streptococcus sp. Lancefield Group G strain.		
XX			
FH	Key	Location/Qualifiers	
FT	Active-site	303..372	
FT		/label=B1	
FT	Active-site	373..427	
FT		/label=B3	
FT	Active-site	443..497	
FT		/label=B2	
XX			
PN	US4956296-A.		
XX			
PD	11-SEP-1990.		
XX			
PF	20-JUN-1988;	88US-0209236.	
XX			
PR	20-JUN-1988;	88US-0209236.	
PR	14-FEB-1986;	86US-0829354.	
PR	23-APR-1986;	86US-0854887.	
PR	17-FEB-1987;	87MO-US00329.	
PR	19-JUN-1987;	87US-0063959.	
XX			
PA	(GENE-) GENEX CORP.		
XX			
PI	Fahnestock SR:		
XX			
DR	WPI: 1990-297491/39.		
DR	N-PsDB: AAO06019.		
XX			
PT	Recombinant Protein G variants - chld. using a cloned gene		
PT	encoding Protein G from Streptococcus sp., used for binding		
PT	immunoglobulin.		
XX			
PS	Disclosure: Fig 9; 48pp; English.		
XX			
CC	Fragments and variants of the sequence are claimed esp. where		
CC	incorporated into a non-pathogenic host eg. E.coli. and expressed		
CC	at high levels.		
CC	The variants have a higher binding efficiency and capacity for		
CC	immunoglobulin, and may be used for purifying, detecting and		
CC	isolating antibodies.		
XX			
SQ	Sequence 593 AA:		
Query Match	24.1%; Score 187;	DB 11;	Length 593;
Best Local Similarity	38.9%; Pred. No. 1.8e-08;		
Matches 61;	Conservative 20;	Mismatches 60;	Indels 16;
Gaps			6;
Oy	1 VDSPEQPIITNGGTLTLLGNAPDEKALRNBERAIDELKKQALIEDKXATTATFAASSD	60	
Db	34 VDSPEDPIITINGGELTNLGNSSETTLALRNBSATADLTMAAVADVAAAAAENAGAA	93	
Oy	61 ALEALADQDALOSEBAAYVKADN--AASDALEAL---ADQTDALOSEBAEYVOSDNA-	114	
Db	94 AEEAAA-AAADALAKKADALEKFNNKYGVSDYIKNLINNAKTYEGVKDLQAQVYESAKRAR	152	
Oy	115 -SDANE-----KAATPIALDVYKTR--DTKEPVKKE	142	



```

Db      153  ISEATDGLSDPLKSGTAPAEVTWKSIETLAEKVLNRE 189
RESULT 16
AAR62944
ID      AAR62944 standard; Protein; 593 AA.
AC      AAR62944;
DT      10-JAN-1995 (first entry)
XX
DE      Streptococcus Protein G derived from strain GX7805.
XX
KM      Streptococcus Protein G; variant; IgG binding activity;
XX      immunoglobulin; Lancefield Group G; bacterial Fc receptor.
XX      Streptococcus sp. GX7805.
XX
FH      Key
FH      Domain
FT      /label= A1
FT      /label= A1
FT      /label= a1
FT      /note= "linking region"
FT      /label= A2
FT      /label= a2
FT      /note= "linking region"
FT      /label= A3
FT      /note= "corresponds to GGA codon"
FT      /label= B1
FT      /note= "involved in IgG binding activity"
FT      /label= b
FT      /note= "linking region"
FT      /label= B3
FT      /note= "involved in IgG binding activity"
FT      /label= B
FT      /note= "linking region"
FT      /label= B2
FT      /note= "involved in IgG binding activity"
FT      /label= C1
FT      /label= C1
FT      /label= C2
FT      /label= C3
FT      /label= C4
FT      /label= C5
FT      /note= "corresponds to GAA codon"
XX
XX      US5312901-A.
XX      17-MAY-1994.
XX
XX      14-FEB-1986; 86US-0829354.
XX      14-FEB-1986; 86US-0829354.
XX

```

```

PR      23-APR-1986; 86US-0854887.
PR      19-JUN-1987; 87US-0063959.
PR      20-JUN-1988; 88US-0209226.
PR      19-JUN-1990; 90US-0540169.
PR      21-APR-1992; 92US-0871539.
XX
PA      (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX
PI      Fahnestock SR;
XX
DR      WPI; 1994-159179/19.
DR      N-PSDB; AA075036.
XX
PT      New recombinant streptococcal protein G variants - useful for
XX      antibody detection and purification and for therapy
XX      Example 5; Fig 9; 48pp; English.
XX
CC      A 2.4Kb HindIII fragment containing the entire coding sequence for
CC      Protein G was isolated from Streptococcus GX7805 using the 1.9kb
CC      Protein G coding sequence from Streptococcus GX7809. The Protein G
CC      has IgG-binding activity which has been localised to the B repeating
CC      structure. Streptococcal Protein G variants comprising the B domains
CC      are claimed.
XX
SQ      Sequence 593 AA;
XX
Query Match 24.1%; Score 187; DB 15; Length 593;
Best local similarity 38.9%; Pred. No. 1.8e-08;
Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;
QY      1 VDSPTEQPRITIPNGGTUNLGNAPKALNBERAIDELKKAIIDEKATTAIEAASSD 60
DB      34 VDSPTEPTPIRNGEGLNLTNGSETTLALRNSATADLTPAAVADTVAAAAENAGAA 93
QY      61 ALAALADQTDALQSEEAAYVRADN--AASDALEAL--ADQTDALQSEEAAYVSDNNA- 114
DB      94 AWEAAL--AADALAKAKADALKEPNKYGVSDYYKNLNNAKTVEGVKDDQAGVDSAKKAR 152
QY      115 -SDAME-----KAATPIALDYKTK--DTKPVYKKE 142
DB      153 ISEATDGLSDPLKSGTAPAEVTWKSIETLAEKVLNRE 189
RESULT 17
AAR10005
ID      AAR10005 standard; protein; 594 AA.
XX
AC      AAR10005;
XX
DT      13-MAR-1991 (first entry)
XX
DE      Streptococcus GX7805 protein G.
XX
KM      Immunoglobulins; Ig.
XX
OS      Streptococcus sp GX7805.
XX
FH      Key
FH      Active-site
FT      /label= Active site B1
FT      /label= Active site B1
FT      /label= Active site B3
FT      /label= Active site B2
FT      /label= Active site B2
XX
XX      US4977247-A.
XX      11-DEC-1990.
XX
XX      19-MAY-1989; 89US-0354264.
XX      19-MAY-1989; 89US-0354264.
XX

```

```

PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87MO-US00329.
PR 19-JUN-1987; 87US-0063959.
PR 20-JUN-1988; 88US-0209236.
XX
XX (GENE-) GENEX CORP.
XX
PI Fehnestock SR, Lee T, Wobble MH:
DR WPI; 1991-006758/01.
DR O-PSDB; 010002.
XX
XX Immobilised protein G variants - used for detection, isolation
PT and purification. Immunoglobulin(s) and immunoglobulin fragments
PS
XX Disclosure; Fig 9; 52pp; English.
XX
XX Protein G gene product may be modified allowing the variant to
CC be immobilised and exhibit different binding profiles. The bound
CC protein is useful in purification and detection of Igs and fragments.
XX
SQ Sequence 594 AA;
Query Match 24.1%; Score 187; DB 12; Length 594;
Best Local Similarity 38.9%; Pred. No. 1.8e-08;
Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;
OY 1 VDSPIDPPIPIRNGGTLNLGNAPKIALRNEERAIDELKKAIEDKKEATTAIEAASD 60
DB 34 VDSPIDPTPIIRNGGELTNLGNSETTLALRNEESATADITAAVADTVAAAAAENAGAA 93
OY 61 ALEALADQTDALQSEEAAYKADN--AASDALEAL--ADQTDALQSEEAAYQSDMAA- 114
DB 94 AMEAAA--AADALAKAKADALKEFNKYGVSQDYKKNLNNAKTVGKVQDIAQAVESAKKAR 152
OY 115 -SDPAWE-----KAATPIALDVKKTK--DTKKYVKKKE 142
DB 153 ISEATDGLSDFLKSGTPAEDTVKSIELAEKVLNRE 189
RESULT 18
AAV57611
ID AAV57611 standard; protein; 103 AA.
XX
XX AAV57611;
XX
XX 10-MAR-2000 (first entry)
XX
DE Streptococcus strain G 148 protein.
XX
XX Streptococcus strain G 148; protein G'; protein G primer; liposome;
KW liposomal delivery complex; connecting moiety; antibody; Fc region;
KW diagnosis; therapeutic agent; antibiotic; antidiarrhoeal; antiviral;
KW antitumorigenic; cytokine; hormone; imaging agent; neurotransmitter;
KW stimulant; cytotoxic agent; malignant cell.
XX
XX Streptococcus sp.
XX
XX WO9959643-A2.
XX
XX 25-NOV-1999.
XX
XX 19-MAY-1999; 99MO-US11177.
XX
XX 20-MAY-1998; 98US-0086347.
XX
XX (SDGS-) SDG INC.
XX
XX Lau JR;
XX
XX WPI; 2000-062383/05.
XX
XX N-PSDB; AA247931.

```

```

XX
XX New liposome constructs comprising a liposome connected to an antibody,
PT used, e.g. for delivery of cytotoxic agents to malignant cells
XX
XX Disclosure; Fig 1; 32pp; English.
XX
XX The present invention describes a construct for connecting an antibody
CC or antibody fragment to a liposome. The construct comprises protein G'
CC and a linking moiety for connecting the protein G' to the liposome.
CC Also described are: (1) a liposomal delivery vehicle comprising: (a) a
CC liposome; and (b) a connecting moiety connected to the liposome, which
CC specifically binds the Fc region of an antibody, for connecting the
CC antibody to the liposome; and (2) forming a liposomal carrier comprising:
CC (a) providing liposomal components having binding moieties, for forming
CC a liposome; (b) providing a construct comprising a linking moiety and a
CC connecting moiety bound together; and (c) combining the liposomal
CC components and the construct and sonifying the combination so that the
CC binding moieties are exposed on a surface of the liposome, for binding to
CC the linking moiety, and to facilitate the binding between them. The
CC products can be used for the delivery of diagnostic or therapeutic
CC agents. The liposomes may contain or may be associated with a diagnostic
CC or therapeutic agent, e.g. antibiotics, antidiarrhoeals,
CC antitumorigenic, antiviral, cytokines, hormones, imaging agents,
CC neurotransmitters, or stimulants. They can be used particularly for the
CC delivery of cytotoxic agents to malignant cells. The protein G'
CC connecting moiety provides a liposomal delivery complex having improved
CC targeting efficiency. As a result of the binding between protein G' and
CC the Fc region of antibodies, protein G' shields the Fc regions of the
CC attached antibodies from non-specific binding to cell-surfaces, other
CC proteins, and anatomical structures. The present sequence represents
CC an unidentified protein encoded by the same sequence which encodes
XX protein G'.
XX
SQ Sequence 103 AA;
Query Match 18.3%; Score 142; DB 21; Length 103;
Best Local Similarity 57.9%; Pred. No. 2.1e-05;
Matches 33; Conservative 3; Mismatches 21; Indels 0; Gaps 0;
OY 1 VDSPIDPPIPIRNGGTLNLGNAPKIALRNEERAIDELKKAIEDKKEATTAIEA 57
DB 34 VDSPIDPTPIIRNGGELTNLGNSETTLALRNEESATADITAAVADTVAAAAAEN 90
RESULT 19
AAP94785
ID AAP94785 standard; protein; 235 AA.
XX
XX AAP94785;
XX
XX 04-JUL-1990 (first entry)
XX
XX Protein G variant.
DE
XX Protein G; immunoglobulin; Fc receptor; ds.
XX
XX Streptococcus sp.
XX
XX WO8810306-A.
XX
XX 29-DEC-1988.
XX
XX 20-JUN-1988; 88MO-US02084.
XX
XX 19-JUN-1987; 87US-0063959.
XX
XX (GENE-) GENEX CORP.
XX
XX Fehnestock SR;
XX
XX WPI; 1989-023848/03.
XX
XX N-PSDB; AA91099.
XX

```

PT Cloned protein G variant genes -  
 PT expressing proteins having immunoglobulin-binding properties of  
 PT protein G and derived from Streptococcus sp.  
 XX  
 PS Claim 17; Page 88; 116pp; English.  
 CC  
 CC Gene for protein G variant of non-pathogenic streptococcus sp. allowing  
 CC isolation of the protein and variants, useful as bacterial Fc receptors  
 CC eg in purification and detection of Abs., screening of hybridoma clones  
 CC and treatment of disease.  
 CC  
 SQ Sequence 235 AA;  
 Query Match 17.9%; Score 139; DB 10; Length 235;  
 Best Local Similarity 25.2%; Pred. No. 0.00011;  
 Matches 58; Conservative 24; Mismatches 64; Indels 84; Gaps 8;  
 OY 4 PIEDPRIIPNGGTLNLTGNAPEKALRNBERAIDE-----LKKQAIEDKEAT 51  
 DB 4 PIEDPTIRNGGELTNLTGSETTLALRNESATAGYPLPKTDYKILNKGKGETT 63  
 OY 52 TAIEAASSD-----ALEALADQTDALQ---SEEAAYKAD-----N 84  
 DB 64 EAVDAATAEKVFKQYANDNGVDGMTYDDATKTFYVEKPEVIDASELPVATYTKLVIN 123  
 OY 85 AASDALALADQTDALQSEEA-EVYOSDNAAADAW----- 118  
 DB 124 GKTLKGETTTKAVDAETAEKAFKQYANDNGVDGVTYDDATKTFYTEMVTEVPRGDAPT 183  
 OY 119 --EK-----AATPIALDVKKTKDT-----KPVYKKEERONVNT 149  
 DB 184 EPEKPEASIPLVLPATPIAKDDAKKDDTKKEDAKKPEAKKDDAKKAET 233  
 RESULT 20  
 AAR07004  
 ID AAR07004 standard; protein; 235 AA.  
 XX  
 AC AAR07004;  
 XX  
 DT 17-JAN-1991 (first entry)  
 XX  
 DE Protein G variant.  
 XX  
 KW Immunoglobulin.  
 XX  
 OS Streptococcus sp. Lancefield Group G strain.  
 XX  
 PN US4956296-A.  
 XX  
 PD 11-SEP-1990.  
 XX  
 PF 20-JUN-1988; 88US-0209236.  
 XX  
 PR 20-JUN-1988; 88US-0209236.  
 PR 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 17-FEB-1987; 87WO-US00329.  
 PR 19-JUN-1987; 87US-0063959.  
 XX  
 PA (GENE-) GENEX CORP.  
 XX  
 PI Fahnestock SR;  
 XX  
 DR WPI; 1990-297491/39.  
 DR N-PSDB; AAC06009.  
 XX  
 PT Recombinant protein G variants - obtd. using a cloned gene  
 PT encoding Protein G from Streptococcus sp., used for binding  
 PT immunoglobulin.  
 XX  
 PS Disclosure; Column 9-16; 48pp; English.  
 XX

CC Sequence may be incorporated into a non-pathogenic host eg. E.coli,  
 CC where they may be expressed at high levels. The proteins have a  
 CC higher binding efficiency and capacity for immunoglobulin, and may  
 CC be used for purifying, detecting and isolating antibodies.  
 CC  
 SQ Sequence 235 AA;  
 Query Match 17.9%; Score 139; DB 11; Length 235;  
 Best Local Similarity 25.2%; Pred. No. 0.00011;  
 Matches 58; Conservative 24; Mismatches 64; Indels 84; Gaps 8;  
 OY 4 PIEDPRIIPNGGTLNLTGNAPEKALRNBERAIDE-----LKKQAIEDKEAT 51  
 DB 4 PIEDPTIRNGGELTNLTGSETTLALRNESATAGYPLPKTDYKILNKGKGETT 63  
 OY 52 TAIEAASSD-----ALEALADQTDALQ---SEEAAYKAD-----N 84  
 DB 64 EAVDAATAEKVFKQYANDNGVDGMTYDDATKTFYVEKPEVIDASELPVATYTKLVIN 123  
 OY 85 AASDALALADQTDALQSEEA-EVYOSDNAAADAW----- 118  
 DB 124 GKTLKGETTTKAVDAETAEKAFKQYANDNGVDGVTYDDATKTFYTEMVTEVPRGDAPT 183  
 OY 119 --EK-----AATPIALDVKKTKDT-----KPVYKKEERONVNT 149  
 DB 184 EPEKPEASIPLVLPATPIAKDDAKKDDTKKEDAKKPEAKKDDAKKAET 233  
 RESULT 21  
 AAR10011  
 ID AAR10011 standard; protein; 265 AA.  
 XX  
 AC AAR10011;  
 XX  
 DT 13-MAR-1991 (first entry)  
 XX  
 DE Type 4 GX7809 protein G variant.  
 XX  
 KW Immunoglobulins; Ig.  
 XX  
 OS Streptococcus sp GX7809.  
 XX  
 PN US4977247-A.  
 XX  
 PD 11-DEC-1990.  
 XX  
 PF 19-MAY-1989; 89US-0354264.  
 XX  
 PR 19-MAY-1989; 89US-0354264.  
 PR 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 17-FEB-1987; 87WO-US00329.  
 PR 19-JUN-1987; 87US-0063959.  
 PR 20-JUN-1988; 88US-0209236.  
 XX  
 PA (GENE-) GENEX CORP.  
 XX  
 PI Fahnestock SR, Lee T, Wroble MH;  
 XX  
 DR WPI; 1991-006758/01.  
 DR Q-PSDB; Q10007.  
 XX  
 PT Immobilised protein G variants - used for detection, isolation  
 PT and purificn. immunoglobulin(s) and immunoglobulin fragments  
 XX  
 PS Disclosure; Column 11; 52pp; English.  
 CC  
 CC Protein G variant product carries active regions B1 and B2. It may  
 CC be immobilised and exhibits different binding profiles. The bound  
 CC protein is useful in purification and detection of Igs and fragments.  
 XX  
 SO Sequence 265 AA;

	Query Match	17.9%	Score 139;	DB 12;	Length 265;	
	Best Local Similarity	25.2%;	Pred. NO. 0.00012;			
	Matches	58;	Conservative	24;	Mismatches	64;
					Indels	84;
					Gaps	8;
QY	4 PLEDPRIIPNGGTLNLGNAPKALNRNEERAIDE-----LKRQAIEDKEAT	51				
Db	34 PLEDPIIRNGELTNLCNSFTLLARNESATAGYPAPKTDYKKLLNCKTLMGETTT	93				
QY	52 TALEAASD-----ALEALADDTDALO-----SEEAAYVKAD-----N	84				
	:   :   :   :   :   :   :   :   :   :   :   :					
Db	94 EAVDAATAEKVPKYQANDNGVDGEITYDATKTFPTVTEKEVIDASELPNAVTKLVIN	153				
QY	85 AASDALREALADDTDALOSEEA-EVVQSDDMAASDAM-----	118				
Db	154 GKTLCGTTRAVDAETAEKAERQYANDNGVGWVTYDATKTFPTVTEKVEVPRDAPT	213				
QY	119 --EK-----AATPIADVKKTKDT-----KPVKKKEQNNTT	149				
Db	214 EPKEEASIPLYPLTPAPPIAKDDKKDDTKEDAKPPAKKDAAKKAAET	263				

RESULT 22  
AAR53294  
ID AAR53294 standard; Protein; 265 AA.

XX	US5312901-A.
PD	
XX	17-MAY-1994.
XX	
PE	14-FEB-1986;
XX	86US-0829354.
PR	14-FEB-1986;
XX	86US-0829354.
PR	23-APR-1986;
XX	86US-0834887.
PR	19-JUN-1987;
XX	87US-0063959.
PR	20-JUN-1988;
XX	88US-0209236.
PR	19-JUN-1990;
XX	90US-0540169.
PR	21-APR-1992;
XX	92US-0871539.
PA	(PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX	
P1	Fahnestock SR;
XX	
DR	WPL; 1994-159179/19.
XX	
DR	N-PADB; AAO64648.
XX	
PT	New recombinant streptococcal protein G variants - useful for
PT	antibody detection and purification and for therapy
XX	
PS	Claim 3; Column 46; 48pp; English.
XX	
XX	Protein G isolated from Streptococcus has IgG-binding activity which
CC	has been localised to the B repeating structure (see AAR53290).
CC	The sequence AAR53294 represents a claimed Streptococcal Protein G
CC	variant comprising the B domain active site and retaining
CC	IgG-binding activity.

xx	Sequence	265 AA;	17.9%;	Score 139;	DB 15;	Length 265;
	Query Match		25.2%;	Pred. No. 0.00012;		
	Best Local Similarity		24;	Mismatches 64;	Indels 84;	Gaps 8;
	Matches					
QY	4	PTEGRRITPNCGLTNLIGNAPEKALRLREERAIDE-----				-----LKQAIEDKEAT 51
DB	34	PEDRPPIIRNGELTNLIGNSETTTLALRNEESAATAGYPLPKDTYKLLINGKTLGGETT				93
QY	52	TAIRAAASD-----ALEALADDTDALD-----				-----SBEAAVVKAD-----N 84
DB	94	EAVDAATFAEKYVKQYANDNGVDGEWTVYDQATFTFVTEKPEVLDASSELPAVATYKLVIN				153
QY	85	AASDALEALADDTDLKQSEEA-EVYQSDMAASDAW-----				-----118
DB	154	GKTLGGETTKAVDAETAEKAKQYANDNGVDGVTYDQATFTFVTEWVTEVPPRGDAPT				213
QY	119	---EK-----AATPIALDVKTKDT-----				-----KPVVKKEEQNVMT 149
DB	214	EPEKKEASIPVLTPTATPIAKDDAKKDDTKEDAKKPEAKDDAKKAEET				263

```

RESULT 23
AAV71038
ID AAV71038 standard; peptide; 28 AA

```

Query Match 16.6%; Score 129; DB 21; Length 28;  
Best Local Similarity 100.0%; Pred. No. 6e-05;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 SDALEALADQTDALQSEAAVVKADNNA 86  
DB 1 SDALEALADQTDALQSEAAVVKADNNA 28

RESULT 24  
AA71036  
ID AA71036 standard; peptide; 23 AA.  
AC AA71036;  
XX  
XX  
XX 29-AUG-2000 (first entry)  
XX  
XX DE Streptococcus pyogenes strain SF370 GRAB protein fragment #1.  
XX  
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KM immune response; Streptococcus pyogenes infection.  
XX  
XX OS Streptococcus pyogenes.  
XX  
XX PN MO200026240-A2.  
XX  
XX PD 11-MAY-2000.  
XX  
XX PF 02-NOV-1999; 99MO-GB03631.  
XX  
XX PR 02-NOV-1998; 98GB-0023975.  
XX  
XX PA (ACT-) ACTINOVA LTD.  
XX  
XX PI Bjorck LH, Rasmussen M;  
XX  
XX DR WPI; 2000-365572/31.  
XX  
XX PT New alpha2M binding protein for generating a protective immune response  
XX to group A streptococcus and purifying the binding protein -  
XX  
XX PS Claim 1; Page 55; 67pp; English.  
XX  
XX CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-56. This fragment is capable of binding alpha2M  
CC and useful in vaccine composition.  
XX  
XX CC  
XX  
XX SQ Sequence 23 AA;

Query Match 15.4%; Score 120; DB 21; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VDSPTEOPRIIPNGTITLNLGN 23  
DB 1 VDSPTEOPRIIPNGTITLNLGN 23

RESULT 25  
AAG38475  
ID AAG38475 standard; Protein; 166 AA.  
AC AAG38475;  
XX

DT 18-OCT-2000 (first entry)  
XX  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 47472.  
XX  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
XX OS Arabidopsis thaliana.  
XX  
XX PN EP1033405-A2.  
XX  
XX PD 06-SEP-2000.  
XX  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0136021.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.

```
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

Query Match 14.9% Score 116; DB 21; Length 166;
Best Local Similarity 26.3%; Pred. No. 0.0078;
Matches 41; Conservative 25; Mismatches 62; Indels 28; Gaps 5;

OY 12 PNCGLTNNLGNP-----EKLALRNEERAIIDLKROATIEDKATTAIEASSDALEALA 66
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 8 PKESDQNEEGSPVKNKPVSENVYAKENNTESGKQJVAETETTSVAKKEFPVPEPK 67
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 67 DQTDALQSE-----EAAVY---KADNAASDALEALADQTDALQSEAEAVYOS 110
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 68 ETPPAVQPEVAAVEESSADAGEAAVVAPEKEVNAATENAEKAEVAVVAPEKEVEY--- 124
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 111 DNAAADAMEKA-ATPIALDVKKTKDKPKPVYKKEERQ 145
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 125 ---AVEAEKKAEPYKAEPEYKAEPEYKKEESKO 157

RESULT 26
AAV71048
ID AAV71048 standard; peptide; 25 AA.
XX
AC AAV71048;
```

XX 29-AUG-2000 (first entry)  
 DT Streptococcus pyogenes GRAB peptide EKL 24.  
 XX  
 DE GRAB: protein G related alpha2m binding protein; vaccine;  
 XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
 KW immune response; Streptococcus pyogenes infection; peptide EKL 24.  
 XX  
 OS Streptococcus pyogenes.  
 XX  
 PN WO200026240-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 02-NOV-1999; 99WO-GB03631.  
 XX  
 PR 02-NOV-1998; 98GB-0023975.  
 XX  
 PA (ACTI-) ACTINOVA LTD.  
 XX  
 PI Bjorck LH, Rasmussen M;  
 XX WPI: 2000-365572/31.  
 DR  
 XX  
 XX New alpha2m binding protein for generating a protective immune response  
 PT to group A streptococcus and purifying the binding protein -  
 XX  
 XX Example 9; Page 31; 67pp; English.  
 PS  
 CC The patent discloses a new family of proteins termed GRAB (protein G  
 CC related alpha2m binding protein) from Streptococcus pyogenes which have  
 CC the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
 CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2m from a sample. The present sequence  
 CC is a GRAB peptide EKL 24 useful in vaccine composition. It was used  
 CC to produce immunoglobulin G (IgG) antibodies specific for native GRAB  
 CC protein in sheep. The peptide has a cysteine insert at the C-terminal  
 CC for attachment to a hetero-bifunctional linker.  
 CC  
 XX  
 SO Sequence 25 AA:  
 Query Match 14.88; Score 115; DB 21; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.00091;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 26 EKILRNERRAIDELKKOAIEDKE 49  
 DB 1 EKILRNERRAIDELKKOAIEDKE 24  
 RESULT 27  
 AAR71928  
 ID AAR71928 standard; Protein; 413 AA.  
 AC AAR71928;  
 DT 22-SEP-1995 (first entry)  
 XX  
 DE S. dysgalactiae MAG protein.  
 XX  
 KW MAG; fast alpha-2-macroglobulin binding protein; FAM;  
 KM plasma proteinase-inhibitor binding protein.  
 XX  
 OS Streptococcus dysgalactiae strain 8215.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..34  
 FT /label- sig.peptide 35..92  
 FT Domain

FT /label- A2M  
 FT /note- "alpha-2-macroglobulin binding domain"  
 FT Domain 93..242  
 FT /label- Alb  
 FT /note- "albumin binding domain"  
 FT Domain 243..312  
 FT /label- IgG  
 FT /note- "IgG binding domain"  
 FT Region 313..387  
 FT /note- "cell wall region"  
 FT Region 388..413  
 FT /label- membrane-spanning region  
 PN WO9507296-A.  
 XX  
 PD 16-MAR-1995.  
 XX  
 PF 06-SEP-1994; 94WO-SE00826.  
 XX  
 PR 06-SEP-1993; 93SE-0002855.  
 XX  
 PA (GUSS/) GUSS B.  
 XX (JONS/) JONSSON H.  
 PA (LIND/) LINDBERG M.  
 PA (MUEL/) MUELLER H.  
 PA (RANT/) RANTAMAKI L K.  
 XX  
 PI Guus B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;  
 XX WPI: 1995-123382/16.  
 DR N-PSDB; AAQ89197.  
 XX  
 PT DNA encoding fast alpha 2-macroglobulin-binding proteins - used  
 PT to obtain prods. for sepn., detection or quantification or for  
 PT binding inhibition  
 PS Disclosure; Fig.4; 50pp; English.  
 XX  
 CC A phage lambda EMBL library of S. dysgalactiae 8215 DNA was analyzed  
 CC for fast-acting alpha-2-macroglobulin (FAM)-, BSA- and IgG-binding  
 CC activity. Clone lambda-SD1, which expressed all 3 activities, was  
 CC analyzed to obtain DNA encoding the FAM-binding protein, MAG. The  
 CC mag gene (given in 089197) encodes a 44 kDa protein (R71929).  
 CC  
 XX  
 SO Sequence 413 AA:  
 Query Match 14.78; Score 114.5; DB 16; Length 413;  
 Best Local Similarity 26.2%; Pred. No. 0.033;  
 Matches 49; Conservative 15; Mismatches 56; Indels 67; Gaps 6;  
 OY 36 AIDELKKOAIED-----KEATTAIE---AASDLEALADOTDLOSEEAAYK 81  
 DB 206 AIEILKRTGIDVYIKLINCKTAEGVALDEILASKPAVIDAPELTPALTYKLYI-- 263  
 OY 82 ADNNAADLEALADOTDLOSEEA-EVQSDNNAASDAWE----- 119  
 DB 264 --NGRTLKGETTTAAVDAETAKAFKOYANANGDGVWYTDATKRTTYTEMTVEPDA 321  
 OY 120 -----KAAPIALDVKKTKDT-----KPYVKEERONVNTLPT 152  
 DB 322 PTERPKKPEASIPVLVFPATPIAKDDAKKDDTKKEDAKKPEAKKPEAKKAAATLPT 381  
 OY 153 TGEESNP 159  
 DB 382 TGEESNP 388  
 RESULT 28  
 AAR71670  
 ID AAR71670 standard; Protein; 413 AA.  
 AC AAR71670;  
 XX

```

DT      21-SEP-1995 (first entry)
XX
XX      S. dysgalactiae MAG protein.
DE
XX      Mag gene; serum albumin binding protein; protein stabilization;
KW      vaccine.
XX
OS      Streptococcus dysgalactiae strain 8215.
XX
XX      Key
FH      Location/Qualifiers
FH      1..34
FH      /label= Sig_peptide
FH      35..92
FH      /label= A2-M
FH      /note= "alpha-2-macroglobulin binding domain"
FH      93..242
FH      /label= Alb
FH      /note= "albumin binding domain"
FH      243..312
FH      /label= IgG
FH      /note= "IgG binding domain"
FH      313..387
FH      /note= "cell wall binding region"
FH      388..413
FH      /note= "membrane-spanning region"
FT      Region
XX
XX      WO9507300-A.
XX
XX      16-MAR-1995.
XX
XX      06-SEP-1994; 94WO-SE00825.
XX      06-SEP-1993; 93SE-0002856.
XX
XX      (GUSG/) Guss B.
PA      (JONS/) JONSSON H.
PA      (LIND/) LINDBERG M.
XX
XX      Guss B, Jonsson H, Lindberg M;
XX
XX      WPI; 1995-123386/16.
DR      N-PSDB; AMQ86080.
XX
XX      DNA encoding a serum albumin binding protein - used to obtain
PT      prods., detection, quantification, protein
PT      stabilisation or vaccine development
XX
XX      Disclosure: Page 29-30; 38pp; English.
XX
CC      A phage lambda EMBL3 library of S. dysgalactiae 8215 DNA was
CC      analyzed for alpha-2-macroglobulin-, BSA- and IgG-binding activity.
CC      Clone lambda-SD1, which expressed all 3 activities, was analyzed to
CC      obtain lambda encoding the SA-binding protein, MAG. The mag gene
CC      (given in Q86080) encoded a 44 kDa MAG protein (R71670).
CC      Recombinant, immobilized MAG was used for serum albumin affinity
CC      purification, detection and assay.
XX
XX      Sequence 413 AA:
SO
    Query Match          14.7%; Score 114.5; DB 16; Length 413;
    Best Local Similarity 26.2%; Pred. No. 0.033;
    Matches 49; Conservative 15; Mismatches 56; Indels 67; Gaps
OY      36 AIDELKKKOAIED-----KEATTAIE--AASSDALEALADQDALQSEEAAYVK 81
Db      206 AIEIILAKKIGIGDYIKKLINNGKTAGCVIALKDIELILASKRAVIDAPELIPALTYYKLVI-- 263
OY      82 ADMAASDALLAALADQDALQSEEA-EVVOASDMAASDAWE----- 119
Db      264 --NCGTLTGCTTTAKVADETAERAKRKQVANENGVDGVWTDATRTFTYEWTEVTPGDA 321
OY      120 -----KAATPIADVTKTKDT-----KPYVKKKEEQNVNTLTFT 152

```

Dd	3732	PTEPKKPPASLPVPLPFPATIAADAKKDQDTKKDDTKKEDAKKEPAKKEAKKAATLPT	361
Qy	153	TGEESEN	159
Dd	382	TGEGSNP	388
RESULT 29			
ID	AAG38476	standard; Protein: 154 AA.	
XX	AAC38476;		
XX	18-OCT-2000	(first entry)	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 47473.		
XX			
KW	Protein identification: signal transduction pathway; metabolic pathway;		
KM	hybridisation assay; genetic mapping; gene expression control; promoter;		
xx	termination sequence.		
xx			
OS	Arabidopsis thaliana.		
PN	EPI033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127452.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132663.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134370.	
PR	14-MAY-1999;	99US-0134371.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	



PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 18-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145085.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-01455913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149920.  
PR 23-AUG-1999; 99US-0149920.  
PR 23-AUG-1999; 99US-0149920.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 14.4%; Score 112; DB 21; Length 154;  
Best Local Similarity 26.5%; Pred. No. 0.016;  
Matches 40; Conservative 25; Mismatches 58; Indels 28; Gaps 5;

Oy 17 LTNLGNAP-----EKLARNBERAIDELKQAIEDKATTAIEAASSDALEALADOTDA 71  
Db 1 MONEGGSVPNKRFSSENVAVAKENNTGESKONQYAEETTESVEAKETFPPEPKETTPA 60  
Oy 72 LQSE-----BAAVV---KADNAASDALEALADQTDALQSEAEVQSDNAAS 115

```
DB      61 VOPEVAANESESSADAGEAAVVAPEKEVENAATENAEKAVEAAVVAPEKVEY-----AV 114
OY      116 DAMEKA-ATP1ALDYKTKTKTPVVKKEBRO 145
DB      115 EAKKKAEPVKAEPVKAEPVKAEPVKEESKQ 145

RESULT 30
AAP70468
ID      AAP70468 standard; Protein: 480 AA.
XX
AC      AAP70468;
XX
DT      21-MAY-1991 (first entry)
XX
DE      Sequence of polypeptide possessing IgG-binding activity of
DE      protein G from Streptococcus G148.
XX
KW      Antibody-binding; IgG; IgA; Immunoglobulin.
XX
OS      Streptococcus G148.
XX
PN      MO8705631-A.
XX
PD      24-SEP-1987.
XX
PF      20-MAR-1987; 87WO-SE00145.
XX
PR      21-MAR-1986; 86SE-0001325.
XX
PA      (PHAA ) PHARMACIA AB.
XX
PA      (GUSS/) GUSS B M.
XX
PI      Guse BM, Lindberg KM, Flock JT, Uhlen CEM;
XX
DR      WPT: 1987-277686/39.
DR      N-PSDB; AAN70757.
XX
PT      New recombinant DNA molecules - for producing proteins with
PT      IgG-binding specificity of protein G or proteins A and G
XX
PS      Example: Fig 2; 39pp; English.
XX
CC      A recombinant DNA molecule containing a nucleotide sequence which
CC      codes for a protein or polypeptide having the same IgG specificity as
CC      protein G from Streptococcus G148 (AAN70757) is claimed. See, for
CC      example, AAN70754, AAN70755 and AAN70756.
XX
SQ      Sequence 480 AA;

Query Match      14.4%; Score 112; DB 8; Length 480;
Best Local Similarity 31.0%; Pred. No. 0.066;
Matches 44; Conservative 12; Mismatches 62; Indels 24; Gaps 5;

OY      41 KKAQIEDKFAATTAIE---AASSDAL--EALADQTDALQSEAAVVKADMAASDALEALA 94
DB      315 KPEVIDASELTPTAVTYKLVNGTKLKGEBTTKAVDAEFAKFAKQYANDNGVGYWTYD 374
OY      95 DQTDALQSEAEVQSDNMAADAMEK-----AATP1ALDYKTKTKDT-----KP 137
DB      375 DATKTFETVTEM-VTEVPGDAPTEPEKPEASIPLVPI/TPTATPIAKDDAKKDDTKKEDAKRP 433
OY      138 VVKKEBROVNTLPTGSESNP 159
DB      434 EAKKEDAKKAETLPTTGESNP 455

RESULT 31
AAG41392
ID      AAG41392 standard; Protein: 191 AA.
XX
AC      AAG41392;
XX
```

```
XX      18-OCT-2000 (first entry)
DT
XX      Arabidopsis thaliana protein fragment SEQ ID NO: 51493.
DE
XX      Protein identification; signal transduction pathway; metabolic pathway;
KW      hybridisation assay; genetic mapping; gene expression control; promoter;
KW      termination sequence.
XX
OS      Arabidopsis thaliana.
XX
PN      EP1033405-A2.
XX
PD      06-SEP-2000.
XX
PF      25-FEB-2000; 2000EP-0301439.
XX
PR      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123180.
PR      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
PR      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
PR      23-APR-1999; 99US-0130510.
PR      28-APR-1999; 99US-0130891.
PR      30-APR-1999; 99US-0131449.
PR      30-APR-1999; 99US-0132048.
PR      04-MAY-1999; 99US-0132407.
PR      05-MAY-1999; 99US-0132484.
PR      06-MAY-1999; 99US-0132485.
PR      06-MAY-1999; 99US-0132486.
PR      07-MAY-1999; 99US-0132487.
PR      11-MAY-1999; 99US-0132863.
PR      14-MAY-1999; 99US-0134218.
PR      14-MAY-1999; 99US-0134219.
PR      14-MAY-1999; 99US-0134221.
PR      18-MAY-1999; 99US-0134370.
PR      19-MAY-1999; 99US-0134768.
PR      20-MAY-1999; 99US-0134941.
PR      21-MAY-1999; 99US-0135124.
PR      24-MAY-1999; 99US-0135353.
PR      25-MAY-1999; 99US-0135629.
PR      27-MAY-1999; 99US-0136021.
PR      28-MAY-1999; 99US-0136392.
PR      01-JUN-1999; 99US-0136782.
PR      03-JUN-1999; 99US-0137222.
PR      04-JUN-1999; 99US-0137528.
PR      07-JUN-1999; 99US-0137502.
PR      08-JUN-1999; 99US-0137724.
PR      10-JUN-1999; 99US-0138094.
PR      14-JUN-1999; 99US-0138540.
PR      14-JUN-1999; 99US-0138847.
PR      16-JUN-1999; 99US-0139119.
PR      16-JUN-1999; 99US-0139452.
PR      17-JUN-1999; 99US-0139453.
PR      18-JUN-1999; 99US-0139492.
PR      18-JUN-1999; 99US-0139454.
PR      18-JUN-1999; 99US-0139455.
PR      18-JUN-1999; 99US-0139456.
PR      18-JUN-1999; 99US-0139457.
PR      18-JUN-1999; 99US-0139458.
PR      18-JUN-1999; 99US-0139459.
PR      18-JUN-1999; 99US-0139460.
PR      18-JUN-1999; 99US-0139461.
PR      18-JUN-1999; 99US-0139462.
PR      18-JUN-1999; 99US-0139463.
```



AC AAC1391;  
XX 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51492.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hydridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 990S-0121825.  
PR 05-MAR-1999; 990S-0123180.  
PR 09-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130440.  
PR 23-APR-1999; 990S-0130510.  
PR 28-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 04-MAY-1999; 990S-0132407.  
PR 05-MAY-1999; 990S-0132484.  
PR 06-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 07-MAY-1999; 990S-0132487.  
PR 11-MAY-1999; 990S-0132863.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 02-JUL-1999; 990S-0142154.  
PR 06-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142920.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144684.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 06-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.  
PR 10-AUG-1999; 990S-0148171.  
PR 11-AUG-1999; 990S-0148319.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148565.  
PR 13-AUG-1999; 990S-0148684.  
PR 16-AUG-1999; 990S-0149368.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 20-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149902.  
PR 23-AUG-1999; 990S-0149930.  
PR 23-AUG-1999; 990S-0150566.

PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160860.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 109; DB 21; Length 230;

Best Local Similarity 27.3%; Pred. No. 0.049; Mismatches 51; Indels 56; Gaps 7;

QY 27 KLALNREBRALDELK-----QAIED-----KEATTAIEAASDLEALADOTDALQ 73  
DB 46 KLEIOWNEKASELKEIEORERDQATIEDLKKEKTKDAKTQLSLEELKIAVFEQAQKAD 105  
QY 74 SE-----AAVYKADNA-----ASDALEALADOT-DALOSEEAQVQSD 111  
DB 106 AEEHARERLNVALESDPRLAIVVKEASAEELTEALRACRDEFLKTEKSEKRETIK 165  
QY 112 NAASDAWEKA-----ATPIALDVKK-----TKDTKPVKKKEERON 146  
DB 166 AATQDALKAQMAQEAETIVDVELKRRKKAASRIIAESKMAKSKTKEVLEKSKPRS 221

RESULT 33  
AAG41390  
ID AAG41390 standard; Protein; 316 AA.

XX AC AAG41390;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 51491.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KM hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KN termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130049.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140685.  
PR 28-JUN-1999; 99US-0140683.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141287.  
PR 07-JUL-1999; 99US-0142055.  
PR 07-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144634.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145813.  
PR 27-JUL-1999; 99US-0145818.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155559.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 109; DB 21; Length 316;  
Best Local Similarity 27.3%; Pred. No. 0.072;  
Matches 48; Conservative 21; Mismatches 51; Indels 56; Gaps 7;

QY 27 KIALRNEERAIIDELKK-----QAIED-----KATTAIEAASDLEALADODIALD 73  
DB 132 KLELQNERKAHSELKEIEORERDHOAIEDLKETKDKAKTQLSLEBELKIAVEAQAEND 191  
QY 74 SEE-----AAVKKADNA-----ASDLEALADOT-DALOSEAEAYQSD 111  
DB 192 ABEHARERLNVANLESDPFSKASABELTETALACRDEETJKTLEMSEREIEDIK 251  
QY 112 NAASDAWEKA-----ATPIALDYKK-----TKOTKPYVKKERON 146  
DB 252 AATODALKRAEMAQEAFTIYVDVELKRRRKAASRIILAESKMKCAKSTFEVLKSPRSS 307  
RESULT 34  
AAG20162

ID AAG20162 standard; Protein: 180 AA.  
XX  
AC AAG20162;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22242.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123160.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125768.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147132.  
PR 05-AUG-1999; 99US-0147260.  
PR 05-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147433.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148317.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.

PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158322.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	14.0%	Score 108.5	DB 21	Length 180
Best Local Similarity	24.0%	Pred. No. 0.04		
Matches	36	Conservative	31	Mismatches 74; Indels 9; Gaps 4
Oy	12	PNGGTLTLLNLCNAP-----EKTLALRNEERAI	DELKQATEDEK	EATTAIBAASDALLEALA 66
Db	8	PKESYMOHEGGSVP	PKPVS	ENVVAKENNTSESEKONQVAETSETTSVAKETSPVEPTPK 67
Oy	67	DOTDALOSEBAVVRKADNNA	SDALLEALA-DOTDALOSEBAE	EVQSSONNAASDAMEKATPI 125
Db	68	ETTPAAEPREVAAVVESSSA	AGAAVAAPKEK	KATENAE-AKVEBAVAAPKEK--VEV 124
Oy	126	ALDYAKTKTDKTPRVYAKKEE	RONNTIPTTGE	155
Db	125	AVEAEKKAEAEFPVKEAE	FPVKA	EAEAEVKA 154
RESULT	35			

ABR62318  
ID ABR62318 standard; Protein; 1096 AA.  
XX  
AC ABR62318;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 13746.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL06421.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 13746; 21pp + Sequence Listing; English.  
XX  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
XX (ABR57737-ABR72072).  
XX  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1096 AA;

```

Query Match Similarity 13.9%: Score 108; DB 22; Length 1096;
Best Local Similarity 21.8%: Pred. No. 0.42; Mismatches 61; Indels 26; Gaps 6
Matches 36; Conservative 42;

QY 11 IPNGCTLNLLGNAPKELALRNE-----ERAIDELKQATIEDKATTALEAAS---SD 60
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 874 VFN-----INSQPKMVDSEAEADKASKAEVQIKKEAEFLNDQEEFTKEDSPIHSD 926
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 61 ALEALADQDIALDQSEAAVYVKADNMAASDLALADQDIALDQDIALDQSEAAVYVS-DNAASDAWE 119
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 927 SVAELREAVTASEGEDVHLEADNITKELLDLDELIAEAER-PDQEKDIVQSEENATTEALD 965
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 130 KAAT-----PIALDVKKTKDTPKPVVKKKEERONVNTLPTTGEESS 157
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 986 RSVTDEDDLVPTQVSTEDQMEIDEPAKAEKAEENEDTFRVADEKEA 1030
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 36
AAG20163
ID AAG20163 standard; Protein: 168 AA.
XX
AC
AG20163;

```



DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 22243.  
DE  
XX Protein Identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 200DEP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145513.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.

```

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153738.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157253.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 13.6%; Score 105.5; DB 21; Length 168;
Best Local Similarity 24.1%; Pred. No. 0.068;
Matches 35; Conservative 31; Mismatches 70; Indels 9; Gaps 4;

```

```

OY 17 LTNLLGNAP-----EKLAIKNEERAIIDELKKQAIEDKEATTATIEAASSDALEALADQDTA 71
DB 1 MONEGSPVKNPKFSENVVAKENNTESGEKONQTVATSETTSVEAKETSPVEPTKETTPA 60
OY 72 LQSEEAAYVKADNNAASDALEALA-DQTDALQSEEAAYVQSDNAASDAWEKAATPTALDYK 130
DB 61 AEPVAAVESSAAAGAAEAAVAAPERKAAATEAEE-AKEVAVAAPERK--VEVAVEAE 117
OY 131 KTKDKPVPVKKKEERONVNTLPTTGE 155
DB 118 KKAEEAPVKVEAEPAKAEAPVKAE 142

```

```

RESULT 37
AAW27286
ID AAW27286 standard; Protein: 314 AA.
XX
AC AAW27286;

```

```

XX 27-APR-1998 (first entry)
DT Rabbit enteropathogenic E. coli (RDEC-1) EspB virulence factor.
DE
XX
DE
XX
KW EspB; rabbit enteropathogenic E. coli; RDEC-1; virulence factor;
KM pathogen; infection; diagnosis; therapy; vaccine; gastroenteritis.
XX
OS Escherichia coli.
XX
PN MO9740063-AZ.
XX
PD 30-OCT-1997.
XX
PF 23-APR-1997; 97WO-CA00265.
XX
PR 23-APR-1996; 96US-0015999.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Finlay B, Kenny B, Stein M;
XX
DR WPL 1997-535772/49.
XX
N-PSDB: AAT91243.
XX
PS
XX
XX
CC This sequence comprises the EspB virulence factor of a rabbit
CC enteropathogenic Escherichia coli (RDEC-1) strain. Its amino acid
CC sequence was deduced from an isolated espB gene (see AAT91293). The
CC predicted EspB polypeptide shows sequence homology more to
CC enterohaemorrhagic E. coli (EHEC) than to enteropathogenic E. coli
CC (EPEC). Claimed novel, isolated EspA polypeptides (see also
CC AAW27284-5) are characterised by: (a) being a secreted protein from
CC EPEC or EHEC; and (b) having a mol.wt. of about 25 kD as determined
CC by SDS-PAGE. Recombinant EspA polypeptides can be expressed in host
CC cells. They can be used to immunise a host susceptible to disease
CC caused by EspA-producing E. coli, and in a method for ameliorating
CC disease caused by EspA-producing organisms. Antibodies raised
CC against EspA can be used in the diagnosis of infection.
XX
SQ Sequence 314 AA;

```

```

Query Match 13.6%; Score 105.5; DB 18; Length 314;
Best Local Similarity 28.7%; Pred. No. 0.15;
Matches 45; Conservative 23; Mismatches 68; Indels 21; Gaps 6;

```

```

OY 14 GGTNLNGLNAPEKIALRNEERAIIDELKKQA--TIDKEATT--AIEAASDALEALADQ 68
DB 100 GCAISSVLA-----ILGSPAIINSATKAGSDIAOKTASTSKAIDAASTDATITLTKA 152
OY 69 T-----DALQSEEAAYVKADNNAASDALEALADQTDALQSEEAAYVQSDNAASDAWEKAATP 124
DB 153 TESVADAVEEDASVQQAQMTTATRAASRTSDVAADIDASQARASQLAEWADADAQAKASHA 212
OY 125 ----IALDYKTKTKDKPVPVKKKEERONVNTLPTTGEES 157
DB 213 SRFMAAVD-KITGSPITAVTSLAGCTKTLPTTVSES 248

```

```

RESULT 38
AAR71127
ID AAR71127 standard; Protein: 208 AA.
XX
AC AAR71127;
XX
DT 11-OCT-1995 (first entry)
XX
DE Synthetic protein G from gene MD1.

```

XX	Synthetic; protein G; IgG; Fc receptors; heavy chain; leukaemia;
KM	Lymphoma; cancer; autoimmune disease.
XX	
OS	Synthetic.
PN	W09506125-A.
XX	
PD	02-MAR-1995.
XX	
PF	23-AUG-1994; 94WO-US09141.
XX	
PR	23-AUG-1993; 93US-0110653.
XX	
PA	(IMMU-) APPLIED IMMUNE SCI INC.
PI	Lee YM, Okarma TB, Talib S;
XX	
DR	WPI: 1995-106854/14.
XX	N-PADB: AAO84886.
PT	New peptide(s) which bind the Fc region of an immunoglobulin -
PT	comprising a nonterminal portion of the amino acid sequence of
PT	Protein A and/or Protein G
XX	
PS	Disclosure: Fig 4A; 61pp; English.
XX	
CC	The sequence is the prod. of the synthetic protein G gene, MD-1. The
CC	proteins produced by MD-1 are capable of specifically binding
CC	the constant region of the heavy chain of IgG in the same way as
CC	neutral Fc receptors. They can be used to analyse the structure and
CC	function of Fc receptors, as well as in antibody production, cell
CC	culture, diagnosis and therapy. They can be used to treat disorders
CC	such as leukemia and lymphoma, cancer and immune disorders.
CC	See also AAR71123-8.
XX	
SO	Sequence 208 AA:
	Query Match 13.5%; Score 105; DB 16; Length 208;
	Best Local Similarity 29.8%; Pred. NO. 0.098;
	Matches 42; Conservative 12; Mismatches 63; Indels 24; Gaps
OY	41 KKQALDEKATTAIER-----ASSDALEALDDQDALOSEEAAYVKNAAASDALEALA 94
	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :
DB	62 KPEVIDASELPFAVTYTKLVIGKTLKETTTTAKVADEAFKAFAKOYANDCGVGWTVD 121
OY	95 DQTDLQSEAVWGSDNADAMKK-----AATPPLADVKKRKDF-----KP 137
	:   :   :   :   :   :   :   :   :   :   :   :
DB	122 DATKTFTVTDEM-VTEPGDAPTEPEKSPISLPLTPPATPIAKDKARKDTRKEDAKRP 180
OY	138 VYKEERQNVNTIPTTGESN 158
	:   :
DB	181 EAKKEDAKKAEITLPTTGCSN 201
RESULT 39	
AAR71929	
ID	AAR71929 standard; Protein; 664 AA.
XX	
AC	AAR71929;
XX	
DT	22-SEP-1995 (first entry)
XX	
DE	S. dysgalactiae MIG.
XX	
KM	MIG; fast alpha-2-macroglobulin binding protein; FAM;
KW	Plasma proteinase-inhibitor binding protein.
XX	
OS	Streptococcus dysgalactiae strain SC1.
XX	
FH	Key Location/Qualifiers
FT	Peptide 1..30
FT	/label= Sfg-peptide

```

FT      Domain                               209..278
FT      /label= IgG1
FT      /note= "IgG binding domain 1"
FT      Domain                               279..348
FT      /label= IgG2
FT      /note= "IgC binding domain 2"
FT      Domain                               349..418
FT      /label= IgG3
FT      /note= "IgC binding domain 3"
FT      Domain                               419..488
FT      /label= IgG4
FT      /note= "IgC binding domain 4"
FT      Domain                               489..558
FT      /label= IgG5
FT      /note= "IgG binding domain 5"
FT      Region                               559..638
FT      /note= "cell wall spanning region"
FT      Peptide                             630..635
FT      /note= "putative wall anchoring motif"
FT      Region                               639..664
FT      /note= "membrane spanning domain"

W09507296-A.
PD      16-MAR-1995.
XX
XX
XX
PF      06-SEP-1994;    94WO-SEU0826.
XX
PR      06-SEP-1993;    93SE-0002855.
XX
XX      (GUSS/) GUSS B.
PA      (JONS/) JONSSON H.
PA      (LIND/) LINDBERG M.
PA      (MUEL/) MUELLER H.
PA      (RANT/) RANTAMAKI L K.
XX
PI      Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;
DR      WPI: 1995-123382/16.
DR      N-PSDB; AAO89196.
XX
PT      DNA encoding fast alpha 2-macro:globulin-binding proteins - used
PT      to obtain prods. for sepn., detection or quantification or for
PT      binding inhibition
XX
PS      Disclosure: Page 31; 50pp; English.
XX
CC      A phage lambda GEM-11 library of S. dysgalactiae SC1 DNA was analyzed
CC      for fast-acting alpha-2-macroglobulin (FAM)- and IgG-binding
CC      activity. One clone, which expressed both activities, was
CC      analyzed to obtain DNA encoding the FAM-binding protein, MIG
CC      (AAR71929). The mig gene is given in AAO89196.
CC
XX      Sequence        664 AA:
XX
Query Match          13.5%; Score 105; DB 16; Length 664;
Best Local Similarity 23.9%; Pred. No. 0.41;
Matches   54; Conservative 21; Mismatches 75; Indels 76; Gaps
OY      1 VDSPIEQP-----RIIPNGGTLTNLGNAPFKALARNBERAIDELKKQAIED----- 47
DB      423 IDAPELTPALTYYKLIVKGNTFES---GETTYTAAV--DAETAAKAFFKOYANENGUYGEWSY 477
OY      48 KRATTAIFAASDDALEALADOTDALOSEAAAYVKADNAASDLEALADOTDALOSEEA-E 106
DB      478 DDATEFTFYTERPAVIDAPELTPALTYYKLV-----NGKTLKGETTTRKAVDAETAERK 533
OY      107 WVOSDNFAASDAW-----EK-----AATPLADV 129
DB      534 OXANENGVGVVTYTDATXTFTYTEVTVPGDAPTEPEKPPASITPLVLTPATPLAKDD 593
OY      130 KKTKQPKPV-----VKKEERONVMTLPPTTGESNP 159

```

DB 594 AKKDKKVVDTKKEDAKKPEAKKEAKKEAKKATLPTTGGSNP 639  
RESULT 40  
AAG37153  
ID AAG37153 standard; Protein: 151 AA.  
XX AAG37153;  
AC  
XX 18-OCT-2000 (first entry)  
DT  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45637.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
XX EPI033405-A2.  
PM  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137502.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142927.  
PR 13-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0143642.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144088.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149902.  
PR 23-AUG-1999; 9905-0149930.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.  
PR 01-SEP-1999; 9905-0151930.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154039.  
PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 28-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158232.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159293.  
PR 13-OCT-1999; 9905-0159294.  
PR 14-OCT-1999; 9905-0159295.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 14-OCT-1999; 9905-0159638.  
PR 18-OCT-1999; 9905-0159584.  
PR 21-OCT-1999; 9905-0160741.  
PR 21-OCT-1999; 9905-0160767.  
PR 21-OCT-1999; 9905-0160768.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160814.  
PR 21-OCT-1999; 9905-0160815.  
PR 22-OCT-1999; 9905-0160980.  
PR 22-OCT-1999; 9905-0160981.  
PR 22-OCT-1999; 9905-0160989.  
PR 25-OCT-1999; 9905-0161404.  
PR 25-OCT-1999; 9905-0161405.  
PR 25-OCT-1999; 9905-0161406.  
PR 26-OCT-1999; 9905-0161359.  
PR 26-OCT-1999; 9905-0161360.  
PR 26-OCT-1999; 9905-0161361.  
PR 28-OCT-1999; 9905-0161920.  
PR 28-OCT-1999; 9905-0161992.  
PR 28-OCT-1999; 9905-0161993.  
PR 29-OCT-1999; 9905-0162142.

Query Match 13.4%; Score 104.5; DB 21; Length 151;

Best Local Similarity 26.0%; Pred. No. 0.073; Mismatches 57; Indels 29; Gaps 6;

DB 12 PNCGLTINLGNAP-----EKLALRNEERAIIDELKKQATIEDKEATTAIEAASSDALPALA 66  
DB 8 PKESDMQNEEGSVNKNVSENVAKENNTESGEKONQTVATTETTSVEAKETFPVPEPTK 67  
QY 67 DQTDALQSE-----EAAVY---KADNAASDALEALADQTDALQSEEAAYQVS 110  
DB 68 ETTTFAVQPEVAEAAVESSADAGEAAVAPKEVNAATEAENAEKAVEAAVAAPEKVEV--- 124  
QY 111 DNAAADAMEKA-ATPIALDYKTK-DTKPV 138

Db 125 ---AVEAEKKRAEAPVKAEEAPVKAEEAPV 151

Search completed: October 13, 2002, 04:45:33  
Job time : 67.6203 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:50:13 ; Search time 330.61 Seconds

(without alignments)  
1143.085 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALFALNDQTDALDSEFAAVKADNAA 28

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ .p2n.model -DEV=slh  
-O/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_093106\_5005/app-query.fasta\_1.526  
-DB=EST -OPMT=fastap -SUFFIX=std.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09847539.ecgn\_1\_1\_763\_etrunat\_10102002\_093106\_5005 -MCP=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:.\*  
1: em\_estdb:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_hic:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	41.9	788	10	BM400696
2	54	41.9	861	10	BM396025

3	53	41.1	613	10	BM012116
4	53	41.1	615	10 <td>BM011078</td>	BM011078
5	53	41.1	624	10 <td>BM013747</td>	BM013747
6	53	41.1	1101	12	CNS005X5
7	52	40.3	438	9	AA941632
8	52	40.3	553	9	BE056346
9	52	40.3	611	12	AG051191
10	52	40.3	1163	12	CNS06M17
11	51.5	39.9	682	10	BM441667
12	51	39.5	243	10	BF496893
13	51	39.5	333	9	AA140863
14	51	39.5	383	9	AI552971
15	51	39.5	387	9	AI467576
16	51	39.5	454	9	AA423358
17	51	39.5	491	9	AI405338
18	51	39.5	529	9	AI0101404
19	51	39.5	535	9	AA540164
20	51	39.5	548	9	AA659168
21	51	39.5	553	10	BM431282
22	51	39.5	558	9	AU206045
23	51	39.5	563	10	BF486089
24	51	39.5	567	9	AI063297
25	51	39.5	567	9	AI388871
26	51	39.5	567	10	BI637415
27	51	39.5	568	9	AI109571
28	51	39.5	569	9	AA392809
29	51	39.5	581	9	AI388559
30	51	39.5	583	9	AI108834
31	51	39.5	615	12	AZ698116
32	51	39.5	622	9	BB201301
33	51	39.5	633	9	AI062828
34	51	39.5	643	9	AA391487
35	51	39.5	650	10	BE337641
36	51	39.5	660	10	BI633216
37	51	39.5	666	9	AI062753
38	51	39.5	666	12	AZ699109
39	51	39.5	710	10	BI637204
40	51	39.5	727	10	BF494445
41	50.5	39.1	561	12	TA2618020
42	50	38.8	445	10	BG440448
43	50	38.8	483	10	BF487636
44	50	38.8	495	9	AU208069
45	50	38.8	528	9	AV393274

#### ALIGNMENTS

RESULT 1  
BM400696  
LOCUS  
DEFINITION  
Tetrahymena thermophila CDNA, mRNA sequence.  
ACCESSION  
BM400696  
VERSION  
BM400696.1 GI:18200749  
KEYWORDS  
EST.  
ORGANISM  
Tetrahymena thermophila.  
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;  
Hymenostomatida; Tetrahymenina; Tetrahymena.  
REFERENCE  
1. (bases 1 to 788)  
Turkewitz R.P., Karer, K.M., Jahn, C., Orlas, E., Kirk, K.E., Frankel  
J. and Klobutcher, L.  
EST from Tetrahymena thermophila, strain CU428.1, growing cells  
Unpublished (2002)  
COMMENT  
Contact: Turkewitz AP  
Molecular Genetics and Cell Biology  
University of Chicago  
920 E. 58th Street, Chicago, IL 60637, USA  
Tel: 773 702 4374  
Fax: 773 702 3172  
Email: apturkew@midway.uchicago.edu  
Seq primer: T3.  
Location/Qualifiers

#### FEATURES

source	1.788	/organism="Tetrahymena thermophila"
		/strain="CU428.1"
		/db_xref="taxon:5911"
		/clone_lib="Chlicoat/Turkewitz cDNA (large fraction)"
		/note="Vector: Bluescript2 SK+, details on library preparation can be found in Chlicoat and Turkewitz (2001) Proc. Natl. Acad. Sci USA, 98: 8709-8713."
BASE COUNT	248 a 219 c 121 g 199 t	1 others
ORIGIN		
Alignment Scores:		
Pred. No.:	103	Length: 788
Score:	54.00	Matches: 12
Percent Similarity:	66.67%	Conservative: 4
Best Local Similarity:	50.00%	Mismatches: 8
Query Match:	41.86%	Indels: 0
DB:	10	Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x BM400696 (1-788)		
OY	2	AspalaLeuGluaLaLeuaLaaspctnrhAspalaLeuGlucGluaLaLaVal 21
DB	265	GATGCTCAACAGCCTCTAACAGCAACACTGATGCTTCATGCTGCCAGAACCTCTT 324
OY	22	ValuysAlaasp 25
DB	325	GCCAATGCTGAA 336
RESULT 2		
BM396025		661 bp mRNA linear EST 17-JAN-2002
LOCUS	5009-0-15-603.t.2	Chlicoat/Turkewitz cDNA (large fraction)
DEFINITION		Tetrahymena thermophila cDNA, mRNA sequence.
ACCESSION	BM396025	
VERSION	BM396025.1	GI:18196078
KEYWORDS	EST.	
SOURCE		Tetrahymena thermophila.
ORGANISM		Tetrahymena thermophila.
REFERENCE		Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
AUTHORS		Hymenostomulida; Tetrahymena; Tetrahymena.
TITLE		Turkewitz A.P., Karer, K.M., Jahn, C., Orlas, E., Kirk, K.E., Frankel, J. and Klobutcher, L.
JOURNAL		EST from Tetrahymena thermophila, strain CU428.1, growing cells
COMMENT		Unpublished (2002)
		Contact: Turkewitz AP
		Molecular Genetics and Cell Biology
		University of Chicago
		920 E. 58th Street, Chicago, IL 60637, USA
		Tel: 773 702 4374
		Fax: 773 702 3172
		Email: apurkew@midway.uchicago.edu
		Seq primer: T3.
FEATURES		
source		Location/Qualifiers
	1..861	/organism="Tetrahymena thermophila"
		/strain="CU428.1"
		/db_xref="taxon:5911"
		/clone_lib="Chlicoat/Turkewitz cDNA (large fraction)"
		/note="Vector: Bluescript2 SK+, details on library preparation can be found in Chlicoat and Turkewitz (2001) Proc. Natl. Acad. Sci USA, 98: 8709-8713."
BASE COUNT	278 a 222 c 151 g 207 t	3 others
ORIGIN		
Alignment Scores:		
Pred. No.:	117	Length: 861
Score:	54.00	Matches: 12
Percent Similarity:	66.67%	Conservative: 4
Best Local Similarity:	50.00%	Mismatches: 8
Query Match:	41.86%	Indels: 0
DB:	10	Gaps: 0

```

US-09-847-539A-6_COPY_59_86 (1-28) x BM396025 (1-861)
OY 2 Aspa1a1eug1ua1a1eua1a1asp1c1n1tr1r1asp1a1a1eug1n1ser1c1ug1ua1a1a1a1 21
Db 501 GATGCTCAACGACCTCTTCAAGACAACAACACTGATGCTCTCAATGCTCCGAGAACGCTCTT 560
OY 22 Val1ys1a1a1asp 25
Db 561 GCCAATGCTGAA 572

RESULT 3
LOCUS Bf012116
DEFINITION Bf012116 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA171D09 5',
ACCESSION Bf012116
VERSION Bf012116.1 GI:17354928
KEYWORDS mRNA sequence.
SOURCE Bf012116.1 GI:17354928
ORGANISM Japanese medaka.
Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorphi; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 613)
AUTHORS Kohara,Y., Shln-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shln-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshln1@genes.nig.ac.jp.
FEATURES
source
1..613
location/Qualifiers
/organism="Oryzias latipes"
/strain="Hg-rR"
/db_xref="taxon:8090"
/clone="MF01SSA171D09"
/clone_11b="MF01SSA cDNA"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
BASE COUNT 134 a 136 c 239 g 104 t
ORIGIN
Alignment Scores:
Pred. No.: 104 Length: 613
Score: 53.00 Matches: 12
Percent Similarity: 60.71% Conservative: 5
Best Local Similarity: 42.86% Mismatches: 11
Query Match: 41.09% Indels: 0
DB: 10 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x Bf012116 (1-613)
OY 1 Ser1a1a1a1eug1ua1a1a1eua1a1asp1c1n1tr1r1asp1a1a1eug1n1ser1c1ug1ua1a1a1a1a 20
Db 8 TCGAGCGCGCGCGAGCGAGCGCGGACGACGAGGAGGCGCTTTAAACGAGAGGAGCGCGCG 67
OY 21 Val1Val1ys1a1a1asp1a1a1a1a1a 28
Db 68 GCTTTAACCGAGGAGCGCGCT 91

RESULT 4
LOCUS Bf011078
DEFINITION Bf011078 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA153f11 5',
ACCESSION Bf011078

```



```

VERSION      B011078.1  GI:17367422
KEYWORDS     EST.
SOURCE       Japanese medaka.
ORGANISM     Oryzias latipes

REFERENCE    Kohara,Y., Shih-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
AUTHORS      Medaka EST Project in Takeda's lab
TITLE        Unpublished (2001)
JOURNAL      Contact: Tadasu Shin-i
COMMENT      National Institute of Genetics
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshin@genes.nig.ac.jp.

FEATURES
source      1..615
            /organism="Oryzias latipes"
            /strain="Hd-r"
            /db_xref="taxon:8090"
            /clone="MF01SSA153F11"
            /sex="mixture of female and male"
            /tissue_type="whole embryo"
            /dev_stage="segmentation stage 20 - 25"

BASE COUNT   135 a 137 c 234 g 109 t
ORIGIN

Alignment Scores:
Pred. No.:    105      Length:    615
Score:        53.00    Matches:    12
Percent Similarity: 60.71% Conservative: 5
Best Local Similarity: 42.86% Mismatches: 11
Query Match:  41.09%  Indels:      0
DB:           10      Gaps:        0

US-09-847-539a-6_COPY_59_86 (1-28) x B011078 (1-615)
OY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
    |||::||| ::| |||||::| ||||| ||||| ||||| |||||
DB 40 TCAGAGCGCGCGGCGAGCGCGGCGGAGGAGCGGCTTTAACCGAGAGAGCGCGCG 99
OY 21 ValValIysAlaAspAsnAlaAla 28
    ::| ::| |||||
DB 100 GCTTTAACCGAGAGGAGCGGCT 123

RESULT 5
B013747 B013747 624 bp mRNA linear EST 05-DEC-2001
LOCUS    B013747 MF01SSA CDNA Oryzias latipes CDNA clone MF01SSA155A09 5',
DEFINITION mRNA sequence.
ACCESSION B013747
VERSION    B013747.1 GI:17363622
KEYWORDS  EST.
SOURCE     Japanese medaka.
ORGANISM  Oryzias latipes
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
          Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
          Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 624)
AUTHORS   Kohara,Y., Shih-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE     Medaka EST Project in Takeda's lab
JOURNAL   Unpublished (2001)
COMMENT   Contact: Tadasu Shin-i
          Center For Genetic Resource Information
          National Institute of Genetics
          1111 Yata, Mishima, Shizuoka 411-8540, Japan
          Tel: 81-559-81-6856

```

```

Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers
1..624
/organism="Oryzias latipes"
/strain="Hd-r"
/db_xref="taxon:8090"
/clone="MF01SSA155A09"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"

BASE COUNT   144 a 139 c 232 g 108 t 1 others
ORIGIN

Alignment Scores:
Pred. No.:    107      Length:    624
Score:        53.00    Matches:    12
Percent Similarity: 60.71% Conservative: 5
Best Local Similarity: 42.86% Mismatches: 11
Query Match:  41.09%  Indels:      0
DB:           10      Gaps:        0

US-09-847-539a-6_COPY_59_86 (1-28) x B013747 (1-624)
OY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
    |||::||| ::| |||||::| ||||| ||||| ||||| |||||
DB 65 TCAGAGCGCGCGGCGAGCGCGGCGGAGGAGGAGGCTTTAACCGAGAGAGCGCGCG 124
OY 21 ValValIysAlaAspAsnAlaAla 28
    ::| ::| |||||
DB 125 GCTTTAACCGAGAGGAGCGGCT 148

RESULT 6
CNS00FX5/C 1101 bp DNA linear GSS 03-JUN-1999
LOCUS      BACR32D16 of RPCI-98 library from Drosophila melanogaster (fruit
DEFINITION fly) genomic survey sequence.
ACCESSION  AL071361
VERSION     AL071361.1 GI:4951201
KEYWORDS   GSS.
SOURCE      fruit fly.
ORGANISM   Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 1101)
AUTHORS    Direct Submission
TITLE       Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage;
JOURNAL     BP 191 91006 EVRY cedex - FRANCE (E-mail: seqret@genoscope.cns.fr
COMMENT     - Web : www.genoscope.cns.fr)
          Determination of this BAC-end sequence was carried out as part of a
          collaboration with the Berkeley Drosophila Genome Project (BDGP).
          The BDGP is constructing a physical map of the Drosophila
          melanogaster genome using these BACs. For further information
          please see http://www.fruitfly.org The BDGP Drosophila
          melanogaster BAC library was prepared by Kazutoyo Oseegawa and
          Aaron Mammosier in Pieter de Jong's laboratory in the Department of
          Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
          NY. The library is named RPCI-98 and was constructed by partial
          EcoRI digestion of Drosophila DNA provided by the BDGP from the
          isogenic strain y2; cn bw sp; the same strain used for the BDGP's
          P1 and EST libraries. A more detailed description of the library
          and how to order individual BAC clones, the entire library, or
          filters for hybridization from the BACPAC Resource Center can be
          found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source      1..1101
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_lib="RPCI-98"

```

/clone="BACR32D16"  
/note="end : 17"  
BASE COUNT 268 a 329 c 226 g 231 t 47 others  
ORIGIN

Alignment Scores:  
Pred. No.: 236 Length: 1101  
Score: 53.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.09% Indels: 0  
DB: 12 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x CNS00FX5 (1-1101)

QY 5 GUAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValAlaAla 24  
||| |||||::: ||||| ::::|||::: ||||| |||  
DB 263 GAGGAGCAAGCTGACGAGGAGGAGCCGCCGAGGAGAGAGATCCGCNANGCTGCCGCC 204

QY 25 ASPAsnAlaAla 28  
||| |||||  
DB 203 GATGCTGGCGGC 192

RESULT 7  
AA941632/c 438 bp mRNA linear EST 19-APR-2001  
LOCUS LD25873.5prlme LD Drosophila melanogaster embryo pot2 Drosophila  
DEFINITION melanogaster cDNA clone LD25873 5prlme, mRNA sequence.  
ACCESSION AA941632  
VERSION AA941632.1 GI:3101545  
KEYWORDS EST.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscimorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 438)  
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
Lewin, S., and Rubin, G.M.  
BDGP/HIMI Drosophila EST Project  
JOURNAL Unpublished (2001)  
COMMENT Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST/estfruitfly.berkeley.edu>  
Plate: 258 row: G column: 1  
High quality sequence stop: 226.  
Location/Qualifiers  
FEATURES  
source 1..438  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="LD25873"  
/clone\_1lb="LD Drosophila melanogaster embryo pot2"  
/sex="male and female"  
/dev\_stage="30 to 24 hours mixed stage embryonic"  
/lab\_host="X11 Blue"  
/note="Organ: embryo. Vector: pOT2. Site\_1: EcoRI; Site\_2:  
XhoI; Sized fractionated cDNAs were directly ligated into  
pOT2."  
BASE COUNT 98 a 125 c 99 g 116 t  
ORIGIN

Alignment Scores:  
Pred. No.: 93.8 Length: 438  
Score: 52.00 Matches: 11  
Percent Similarity: 55.56% Conservative: 4  
Best Local Similarity: 40.74% Mismatches: 12  
Query Match: 40.31% Indels: 0  
DB: 9 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AA941632 (1-438)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
:::|||||::: ||| |||||::: ||| |||||  
DB 231 AACGATCGCGCGCAATCATCTATACAGACCCACAGACGAGAGAGACAGACAGCAGC 172

QY 21 ValValLyAlaAspAsnAla 27  
|||::: ||| |||  
DB 171 CGGTACGAGGAGATCGTCC 151

RESULT 8  
BE056346/c 553 bp mRNA linear EST 08-JUN-2000  
LOCUS 00257 leafy spurge Lambda HybridZAP 2.1 two-hybrid vector cDNA  
DEFINITION library Euphorbia esula cDNA clone 25D 5' similar to Arabidopsis  
unknown Protein (accession# AF24521), mRNA sequence.  
ACCESSION BE056346  
VERSION BE056346.1 GI:8383403  
KEYWORDS EST.  
SOURCE leafy spurge.  
ORGANISM Euphorbia esula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Malpighiales; Euphorbiaceae; Euphorbia.  
REFERENCE 1 (bases 1 to 553)  
Anderson, J.V. and Horvath, D.P.  
Identification of mRNAs expressed in underground adventitious buds  
of Euphorbia esula (leafy spurge)  
JOURNAL Unpublished (2000)  
COMMENT Contact: Anderson JV  
Plants Science Research  
USDA/ARS, Biosciences Research Lab  
1605 Albrecht Blvd., PO Box 5674, Fargo, ND 58105, USA  
Tel: 701 239 1263  
Fax: 701 239 1252  
Email: [andersjv@fargo.ars.usda.gov](mailto:andersjv@fargo.ars.usda.gov)  
Seq primer: PAD5.  
Location/Qualifiers  
FEATURES  
source 1..553  
/organism="Euphorbia esula"  
/db\_xref="taxon:3993"  
/clone="25D"  
/clone\_1lb="leafy spurge Lambda HybridZAP 2.1 two-hybrid  
vector cDNA library"  
/tissue\_type="underground adventitious buds"  
/dev\_stage="3-day induced (decapitated)"  
BASE COUNT 161 a 117 c 116 g 159 t  
ORIGIN

Alignment Scores:  
Pred. No.: 130 Length: 553  
Score: 52.00 Matches: 12  
Percent Similarity: 65.38% Conservative: 5  
Best Local Similarity: 46.15% Mismatches: 9  
Query Match: 40.31% Indels: 0  
DB: 9 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x BE056346 (1-553)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
|||:::|||||::: ||| |||||::: ||| |||||  
DB 541 TCAGAGCACCTTCTTCACTGATGCACAAAGCGATGATCAGCAGCAGCCTTGCC 482

QY 21 ValValLyAlaAspAsn 26  
||| |||||  
DB 481 ATTTCATACACGATATAT 464

RESULT 9  
AO951191/c 611 bp DNA linear GSS 27-JAN-2000  
LOCUS AO951191  
DEFINITION Sheared DNA-50A17.7R Sheared DNA trypanosoma brucei genomic clone  
Sheared DNA-50A17, DNA sequence.  
ACCESSION AO951191  
VERSION AO951191.1 GI:6774456

**KEYWORDS** GSS.

**SOURCE** Trypanosoma brucei.

**ORGANISM** Trypanosoma brucei  
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;

**REFERENCE**

**AUTHORS** El-Sayed N., Zhao S., Zhao H., Gill S., Suh E., Malek J., Fujii C., Gerrard C., Leech V., de Jong P., Ullu E., Melville S., Donelson J.

**TITLE** I (bases 1 to 611)

**JOURNAL** Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library

**COMMENT** Unpublished (1999)  
Other GSSs: Sheared DNA-50A17.TF  
Contact: Majib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tigr.org  
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page:  
<http://www.tigr.org/tdb/mdb/vbdb/>.  
Seq primer: M13-Reverse  
Class: Shotgun.  
Location/Qualifiers  
1..611  
/organism="Trypanosoma brucei"  
/strain="TREN927/4 GUTat 10.1"  
/db\_xref="taxon:5691"  
/clone="Sheared DNA-50A17"  
/note="Vector: pUC18; Site\_1: SmaI; Constructed at The Institute for genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREN927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaundin and B. Barrell, Oxford University Press, 1999)." .

**BASE COUNT** 170 a 175 c 96 g 170 t

**ORIGIN**

**Alignment Scores:**

Pred. No.: 149 Length: 611

Score: 52.00 Matches: 13

Percent Similarity: 61.54% Conservative: 3

Best Local Similarity: 50.00% Mismatches: 6

Query Match: 40.31% Indels: 4

DB: 12 Gaps: 1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AQ951191 (1-611)

OY 7 LeuAlaASPGLThrAspAla-----LeuGlnSerGlutGluAlaAlaValVal 22  
||| |||||| :||| ||| :||| ||||||| :

DB 596 CTAAAGCATCAAGGTGGAGCGCACACACACTTTGTTGGCGACGTCGACGATGCCA 537  
||| |||||||||

OY 23 LysAlaASPAsnAlaAla 28  
||| |||||||||

DB 536 AAGGGTGATTAATGCAGCC 519

**RESULT 10**

CNS06N17 1163 bp DNA linear GSS 17-JUN-2001

LOCUS T3 end of clone AU0A012D04 of library AU0A from strain CBS 3082

DEFINITION of Saccharomyces kluyveri, genomic survey sequence.

ACCESSION AL406805

VERSION AL406805.1 GI:12171287

**KEYWORDS** GSS.

VERSION BM441667.1 GI:18472442  
 KEYWORDS  
 SOURCE barley  
 ORGANISM Hordeum vulgare  
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae ; Triliceae; Hordeum.  
 1 (bases 1 to 682)  
 REFERENCE Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L., Ramsay, L., Machray, G., Marshall, D.F.M. and Maugh, R.  
 TITLE Development of Barley Transcriptome Resources  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Maugh R  
 Unit of Genomics  
 Scottish Crop Research Institute  
 Invergowrie, Dundee, DD2 5DA, Scotland, UK  
 Tel: 00 44 1382 562731  
 Fax: 00 44 1382 562426  
 Email: rmaugh@scri.scri.ac.uk  
 All sequence has a phred quality score of 20 or over  
 Seq primer: M13 reverse.

FEATURES  
 source  
 Location/Qualifiers  
 1..682  
 /organism="Hordeum vulgare"  
 /cultivar="Optic"  
 /db\_xref="taxon:4513"  
 /clone="EBed07\_S0001.D10"  
 /clone\_lib="IGF Barley EBed07 library"  
 /library\_type="Endosperm"  
 /dev\_stage="28 days post anthesis"  
 /lab\_host="DH10B"  
 /note="Vector: pSPORT1; Site\_1: Sal I; Site\_2: Not I; Non-normalised library, directionally cloned into pSPORT1. Derived from endosperm tissue dissected from developing grains (28 days post anthesis) in glasshouse grown barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project."

BASE COUNT 117 a 241 c 211 g 113 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 208 Length: 682  
 Score: 51.50 Matches: 17  
 Percent Similarity: 67.86% Conservative: 2  
 Best Local Similarity: 60.71% Mismatches: 8  
 Query Match: 39.92% Indels: 1  
 DB: 10 Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BM441667 (1-682)  
 QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaVal 20  
 ||| ::| ||||| ||| ||||| |||||::| |||||  
 Db 427 TCCTCGTCGAGAGAGCGCTTGCCCGCATGCAGACGCCCTG--ACCGAGATCGCCGC 371

QY 21 ValValLysAlaAspAsnAlaVal 28  
 ||||| ||| |||||  
 Db 370 GTCGCTCTCTGATGATGCGACGC 347

RESULT 12  
 BF496493/c 243 bp mRNA linear EST 19-APR-2001  
 LOCUS A110425.Sprime AT Drosophila melanogaster adult testes POTB7.  
 DEFINITION Drosophila melanogaster cDNA clone A110425 5 similar to CG7289;  
 Fban0007289 located on: 2L 22B8-22B8; 04/08/2001, mRNA sequence.  
 ACCESSION BF496493  
 VERSION BF496493.2 GI:13694046  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 243)  
 AUTHORS Stapleton, M., Brockstein, P., Hong, L., Agdayan, A., Baxter, E., Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guatin, H., Harris, N., Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nino, J., Paclob, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Cealiker, S. and Rubin, G.M.  
 TITLE BDGP/HMI AT Drosophila EST Project  
 JOURNAL Unpublished (2000)  
 COMMENT On Dec 6, 2000 this sequence version replaced gi:11579867.  
 Contact: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)  
 Plate: AT.104 row: C column: 1  
 High quality sequence stop: 235.

FEATURES  
 source  
 Location/Qualifiers  
 1..243  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="A110425"  
 /clone\_lib="AT Drosophila melanogaster adult testes POTB7"  
 /sex="male"  
 /dev\_stage="0-3 day old Ore-R males"  
 /lab\_host="Plates AT.10-AT.120: DHS-alpha. Plates AT.121-AT.319: DHS-alpha TONA"  
 /note="Organ: ADULT testes; Vector: POTB7; Site\_1: EcoRI; Site\_2: XhoI; The mRNA for the testis library was made from testes and seminal vesicles hand dissected from 0-3 day old Ore-R males. RNA kindly provided by the lab of Margaret Fuller. Sized fractionated cDNAs were directly ligated into POTB7. Plasmid cDNA library."

BASE COUNT 53 a 65 c 56 g 68 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 59.3 Length: 243  
 Score: 51.00 Matches: 11  
 Percent Similarity: 53.85% Conservative: 3  
 Best Local Similarity: 42.31% Mismatches: 12  
 Query Match: 39.53% Indels: 0  
 DB: 10 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BF496493 (1-243)  
 QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaVal 21  
 ||| ::| ||||| ||| ||||| |||||  
 Db 228 GATCGCGGATCATATGAGACGACGACGAGAGACGAGAGGAGGAGGAGGAGGAGGAGG 169

QY 22 ValLysAlaAspAsnAlaVal 27  
 ||||| ||| |||||  
 Db 168 GTACGAGGAGATCGTGC 151

RESULT 13  
 A1104863 333 bp mRNA linear EST 29-NOV-1998  
 LOCUS CK00525.Sprime CK Drosophila melanogaster embryo blueScript.  
 DEFINITION Drosophila melanogaster cDNA clone CK00525 Sprime, mRNA sequence.  
 ACCESSION A1104863  
 VERSION A1104863.1 GI:1704336  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE Kopczyński, C., Serano, T., Rubin, G. and Goodman, C.  
 AUTHORS BDGP/HMI CK Drosophila EST Project  
 TITLE Unpublished (1996)  
 JOURNAL Other\_ESTs: CK00525.Sprime



Alignment Scores:	
Pred. No.:	113
Score:	51.00
Percent Similarity:	68.18%
Best Local Similarity:	50.00%
Query Match:	39.53%
DB:	9
	Gaps:
	0

RESULT	16
AA423358/c	
LOCUS	
DEFINITION	454 bp mRNA linear EST 16-OCT-1997
IMAGE:820441.5'	mRNA Mus musculus CDNA clone
ACCESSION	AA423358
VERSION	AA423358.1 GI:2102174
KEYWORDS	EST.
SOURCE	house mouse.

**FEATURES**  
source  
Email: mousestewatson.wustl.edu  
This clone is available royalty-free through INLNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:#88721  
Seq primer: -26ml3 rev2 ET from Amersham.  
Location/Qualifiers  
1..454

Alignment Scores:	
Pred. No.:	142
Score:	51.00
Percent Similarity:	68.18%
Best Local Similarity:	50.00%
Query Match:	39.53%
DB:	9
US-09-847-539A-6_COPY_59_86 (1-28) x AA433358 (1-454)	
	Length: 454
	Matches: 11
	Conservative: 4
	Mismatches: 7
	Indels: 0
	Gaps: 0

RESULT 17	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE
AI405338/c	AI405338	491 bp mRNA linear EST 19-APR-2001	GH25307.5	fruit fly.	
			melanogaster		
			cdna clone		
			GI:4248425		
			EST.		
			fruit fly.		

```

FEATURES
  source
    1..491
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone="GH25307"
      Email: http://www.fruitfly.org/EST\_est@fruitfly.berkeley.edu
      hit genomic sequence AC005749
      Plate: 253 row: A column: 7
      High quality sequence stop: 465.
      Location/Qualifiers

```



LOCUS	AM659168/c	548 bp	mRNA	linear	EST 25-APR-2001
DEFINITION	96120 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.				
ACCESSION	AM659168				
VERSION	AM659168.1	GI:7424995			
KEYWORDS	EST.				
SOURCE	cow.				
ORGANISM	Bos taurus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.				
REFERENCE	1 (bases 1 to 548) Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,T., Cho,J., Fahrenkrug,S.C., Bennett ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-Mckown,C.G., Petlee,G., Holt,I., Karamycheva,S., Langolf, F., Quickendush,J., and Keefe,J.W.				
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle				
JOURNAL	Genome Res. 11 (4), 626-630 (2001)				
MEDLINE	21180013				
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smithhemill.marc.usda.gov Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -mnscore 18 and -mismatch 12 options. PCR primers FORWARD: AGCAACAGCTATGACCAT BACKWARD: GTTTCCTCCAGTCACGACG Plate: 83 fow: M column: 13 Seq primer: ATTGAGTGACACTATAG. Location/Qualifiers 1..548 /organism="Bos taurus" /db_xref="taxon:9913" /clone_id="MARC 1BOV" /tissue_type="pooled" /lab_host="DH10B" /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."				
FEATURES					
SOURCE					
BASE COUNT	132 a 132 c 151 g 122 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	184	Length:	548		
Score:	51.00	Matches:	11		
Percent Similarity:	60.87%	Conservative:	3		
Best Local Similarity:	47.83%	Mismatches:	9		
Query Match:	39.53%	Indels:	0		
	9	Gaps:	0		
US-09-847-539A-6_COPY_59_86 (1-28) x	AM659168 (1-548)				
OY	4	leugluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys	23		
LOCUS	BM431282/c	553 bp	mRNA	linear	EST 31-JAN-2002
DEFINITION	1DU012E08 Bos taurus Duodenum #1 library Bos taurus cDNA, mRNA sequence.				
ACCESSION	BM431282				
VERSION	BM431282.1	GI:18453004			

	KEYWORDS	EST.
SOURCE	COW.	Bos taurus
ORGANISM	Eukaryote:	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
		Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
REFERENCE	AUTHORS	Hansen,C., Fu,A., Meng,Y., Li,C., Okine,E., Sensen,C.W., Gordon,P.M.N. and Moore,S.S. Gene Expression Profiling of the Bovine Gastrointestinal Tract Unpublished (2002) Contact: Dr. Stephen Moore . Beef Genomics Laboratory Dept of AFNS, University of Alberta 410 Agri/Fort. Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada Tel.: 780 492 0169 Fax: 780 492 4265 Email: smoores@afns.ualberta.ca Insert Length: 553 Std Error: 0.00 POLYA-NO.
FEATURES	source	Location/Qualifiers 1..553 /organism="Bos taurus" /db_xref="taxon:9913" /clone_lib="Bos taurus Duodenum #1 library" /tissue_type="Smooth muscle" /cell_type="Simple columnar epithelial" /dev_stage="Young adult" /lab_host="XLI-BIUMRF strain" /note="Organ: Intestine/Duodenum; Vector: Uni-ZAPXR; Site_1: EcoRI; Site_2: Xho I"
BASE COUNT	158 a	124 c
ORIGIN	139 g	132 t
Alignment Scores: Pred. No.:	186	Length: 553
Score:	51.00	Matches: 11
Percent Similarity:	60.87%	Conservative: 3
Best Local Similarity:	47.83%	Mismatches: 9
Query Match:	39.53%	Indels: 0
DB:	10	Gaps: 0
US-09-847-539A-6_COPY_-59_-86 (1-28) x BM4J31262 (1-553)		
OY	4	LengUAlALeuaLaASpGLnThrsPaLLeugInSerGlUGlUAIAlaVaValLyS 23
I I	II	III  IIII::          III  I::
Dd	142	CTTCTCCTCCCTTCGCCAGCGATGCACGGCAAGAAGAGACGCCACCGTGAG 83
OY	24	Alaspaasn 26
I I::	:	: :
Db	82	GAGCATGCAC 74
RESULT 22		
LOCUS	AUZ06045	558 bp mRNA linear EST 17-JUL-2001
DEFINITION	AUZ06045	unpublished oligo-capped cDNA library, stage L1
ACCSSION	AUZ06045	Caenorhabditis elegans CDNA clone yf858ell 5'', mRNA sequence.
VERSION	AUZ06045	GI:14838902
KEYWORDS	EST.	Caenorhabditis elegans.
SOURCE	Caenorhabditis elegans.	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdlitida; Rhabdiloidea
ORGANISM	Rhabdlitids	Peloderinae; Caenorhabdlits.
REFERENCE	kohara,y., Shin-i,t., Thillery-Mleg,j., Thillery-Mleg,D., Suzuki,y.	
AUTHORS	and Sugano,s.	
TITLE	A complementary view of the C.elegans genome	
JOURNAL	Unpublished (2001)	
COMMENT	Contact: Yuji Kohara	
	Genome Biology Lab.	
	National Institute of Genetics	
	yata iihi, Mishima, Shizuka 41j, Japan	



Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.

## FEATURES

Location/Qualifiers

1..558  
/organism="Caenorhabditis elegans"

/strain="N2"

/dev\_stage="L1"

/clone="yk858ell"

/clone\_lib="unpublished oligo-capped cDNA library, stage L1"

/sex="Hermaphrodite"

/tissue\_type="whole animal"

/dev\_stage="L1"

BASE COUNT 164 a 146 c 142 g 105 t 1 others

## ORIGIN

## Alignment Scores:

Pred. No.:	Length:	Matches:	Score:
189	558	9	51.00
Percent Similarity:	Conservative:	8	68.00%
Best Local Similarity:	Mismatches:	8	36.00%
Query Match:	Indels:	0	39.53%
DB:	Gaps:	0	

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AU206045 (1-558)

OY 1 SeraspalaleuglualaleuAlaaspGlnThrAspalaleugGlnSerGluGluAlaAla 20

Db 236 TCACATGCTATTCAAGAGCTTCGATCAATTCAGCAACATTCACAAAGCAAAAGGAGAGA 295

OY 21 ValValysAlaAsp 25

Db 296 ATTGAGAGAGAGAA 310

RESULT 23

BF486089/c

LOCUS

DEFINITION

AT20267.5prime AT Drosophila melanogaster adult testes POTB7

Phan0007289 located on: 2L 22B8-22B8; 04/09/2001, mRNA sequence.

ACCESSION

BF486089

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 563)

REFERENCE

AUTHORS

Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman

, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan

, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harrel, N., Li, P.,

Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nuno, J., Pacle, J.,

Paragas, V., Park, S., Phouenavong, S., Wan, K., Yu, C., Lewis, S.E.,

Celniker, S., and Rubin, G.M.

BDGP/HMT AT Drosophila EST Project

Unpublished (2000)

On Dec 6, 2000 this sequence version replaced gi:11569390.

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic AE003584: arm:2L [1824960,2149443]

estimated-cyto:22B4-22D2: 04/09/2001

Plate: AT 202 row: F column: 7

High quality sequence stop: 562.

Location/Qualifiers

1..563

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT 127 a 161 c 141 g 138 t

/clone="AT20267"

/clone\_lib="AT Drosophila melanogaster adult testes POTB7"

/sex="male"

/dev\_stage="0-3 day old Ore-R males"

/lab\_host="Plates AT-10-AT-120; DH5-alpha. Plates

AT-121-AT-319; DH5-alpha TONA"

/note="Organ: ADULT testes; Vector: POTB7; Site:1: EcoRI;

Site:2: XhoI. The mRNA for the testis library was made

from testes and seminal vesicles hand dissected from 0-3

day old Ore-R males. RNA kindly provided by the lab of

Margaret Fuller. Sized fractionated cDNAs were directly

ligated into POTB7. Plasmid cDNA library."

BASE COUNT 126 a 159 c 141 g 137 t

## ORIGIN

## Alignment Scores:

Pred. No.:	Length:	Matches:	Score:
191	563	11	51.00
Percent Similarity:	Conservative:	3	53.85%
Best Local Similarity:	Mismatches:	12	42.31%
Query Match:	Indels:	0	39.53%
DB:	Gaps:	0	

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BF486089 (1-563)

OY 2 AspalaleuglualaleuAlaaspGlnThrAspalaleugGlnSerGluGluAlaAla 21

Db 228 GATCGCGGATCATATGAGACACACAGCAGAGAGACAGAGAGAGAGAGAGAGAGAG 169

OY 22 ValValysAlaAspAsnAla 27

Db 168 GTACGAGAGATCGTCC 151

RESULT 24

AT063297/c

LOCUS

DEFINITION

GH03030.5prime GH Drosophila melanogaster head POT2 Drosophila

melanogaster cDNA clone GH03030 5prime, mRNA sequence.

ACCESSION

AT063297

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 567)

REFERENCE

AUTHORS

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S., and Rubin, G.M.

BDGP/HMT Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic sequence DS06378

Plate: 30 row: C column: 6

High quality sequence stop: 454.

Location/Qualifiers

1..567

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="GH03030"

/clone\_lib="GH Drosophila melanogaster head POT2"

/sex="male and female"

/dev\_stage="adult"

/lab\_host="DH5-alpha"

/note="Organ: head; Vector: POT2; Site:1: EcoRI; Site:2:

XhoI. Sized fractionated cDNAs were directly ligated into

POT2. Plasmid cDNA library."

BASE COUNT 127 a 161 c 141 g 138 t

```

ORIGIN
Alignment Scores:
Pred. No.: 193
Score: 51.00
Percent Similarity: 53.85%
Best Local Similarity: 42.31%
Query Match: 39.53%
DB: 9
gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x A1063297 (1-567)
Qy 2 AspalaleuGlualaleuAlaAspGlnThrAspalaleuGlnserGIuAlaAlaVal 21
Db 228 GATCGCGCGAATCATATAGAAAGACACAGAGAGAGAGAGAGAAACAGCGCG 169
Qy 22 VallysAlaAspAsnAla 27
Db 168 GTACGAGAGATCTGCTGCC 151

RESULT 25
A1388871/c 567 bp mRNA linear EST 19-APR-2001
LOCUS GH19913.5prtime GH Drosophila melanogaster head pot2 Drosophila
DEFINITION melanogaster cDNA clone GH19913 5prtime, mRNA sequence.
ACCESSION A1388871
VERSION A1388871.1 GI:4202882
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 567)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HM1 Drosophila EST project
Unpublished (2001)
Contact: Stapleton, M.

FEATURES
source
1..567
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD Drosophila melanogaster head pot2"
/dev_stage="adult"
/lab_host="DHS - alpha"
/notes="Organ: head; Vector: pot2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."
Location/Qualifiers
1..567
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD Drosophila melanogaster head pot2"
/dev_stage="adult"
/lab_host="DHS - alpha"
/notes="Organ: head; Vector: pot2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."

BASE COUNT 127 a 161 c 141 g 138 t
ORIGIN
Alignment Scores:
Pred. No.: 193
Score: 51.00
Percent Similarity: 53.85%
Best Local Similarity: 42.31%
Query Match: 39.53%
DB: 9
gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x A1388871 (1-567)
Qy 2 AspalaleuGlualaleuAlaAspGlnThrAspalaleuGlnserGIuAlaAlaVal 21

```

```

Db 227 GATCGCGCGAATCATATAGAAAGACACAGAGAGAGAGAGAGAAACAGCGCG 168
Qy 22 VallysAlaAspAsnAla 27
Db 167 GTACGAGAGATCTGCTGCC 150

RESULT 26
B1637415/c 567 bp mRNA linear EST 10-SEP-2001
LOCUS SD19430.5prtime SD Drosophila melanogaster Schneider L2 cell culture
DEFINITION pot2 Drosophila melanogaster cDNA clone SD19430.5 similar to
CG7289: FBan0007289 GO:[] located on: 2L 22B8-22B8:: 05/19/2001,
mRNA sequence.
ACCESSION B1637415
VERSION B1637415.1 GI:15539625
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 567)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HM1 Drosophila EST project
Unpublished (2001)
Contact: Stapleton, M.

FEATURES
source
1..567
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DHS - alpha"
/notes="Vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."
Location/Qualifiers
1..567
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DHS - alpha"
/notes="Vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."

BASE COUNT 127 a 161 c 141 g 138 t
ORIGIN
Alignment Scores:
Pred. No.: 193
Score: 51.00
Percent Similarity: 53.85%
Best Local Similarity: 42.31%
Query Match: 39.53%
DB: 10
gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x B1637415 (1-567)
Qy 2 AspalaleuGlualaleuAlaAspGlnThrAspalaleuGlnserGIuAlaAlaVal 21
Db 227 GATCGCGCGAATCATATAGAAAGACACAGAGAGAGAGAGAGAAACAGCGCG 166
Qy 22 VallysAlaAspAsnAla 27
Db 167 GTACGAGAGATCTGCTGCC 150

RESULT 27
A1109571/c 568 bp mRNA linear EST 19-APR-2001
LOCUS GH08804.5prtime GH Drosophila melanogaster head pot2 Drosophila
DEFINITION

```

melanogaster cDNA clone GH08804 5prime, mRNA sequence.  
 A1109571 GI:3477895  
 VERSION A1109571.1  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 568)  
 REFERENCE Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
 Lewis, S. and Rubin, G. M.  
 BDGP/HHMI Drosophila EST Project  
 TITLE Unpublished (2001)  
 JOURNAL Contact: Stapleton, M.  
 COMMENT BDGP

FEATURES  
 source  
 Location/Qualifiers  
 1..568  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_1db="GH Drosophila melanogaster head port2"  
 /sex="male and female"  
 /dev\_stage="adult"  
 /lab\_host="DHS - alpha"  
 /note="Organ: head; Vector: port2; Site\_1: EcoRI; Site\_2: XhoI; Sized fractionated cDNAs were directly ligated into port2. Plasmid cDNA library."  
 BASE COUNT 127 a 161 c 142 g 138 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 193 Length: 568  
 Score: 51.00 Matches: 11  
 Percent Similarity: 53.85% Conservative: 3  
 Best Local Similarity: 42.31% Mismatches: 12  
 Query Match: 39.53% Indels: 0  
 DB: 9 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x A1109571 (1-568)  
 Oy 2 AspalaleuGUALaleuAlaaspGlnthraspalaleuGinsersGluAlaAlaVal 21  
 Db 228 GATCGCGGAATCATATAGAACACACAGACGACGAGACGAGAACCAACGCGC 169

RESULT 29  
 LOCUS A1388559/c 581 bp mRNA linear EST 19-APR-2001  
 DEFINITION GH19508 5prime GH Drosophila melanogaster head port2 Drosophila  
 melanogaster cDNA clone GH19508 5prime, mRNA sequence.  
 VERSION A1388559  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 581)  
 REFERENCE Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
 Lewis, S. and Rubin, G. M.  
 BDGP/HHMI Drosophila EST Project  
 TITLE Unpublished (2001)  
 JOURNAL Contact: Stapleton, M.  
 COMMENT BDGP

BDGP/HHMI Drosophila EST Project  
 JOURNAL Unpublished (2001)  
 COMMENT On Jan 19, 1998 this sequence version replaced g1:2792678.  
 Other\_ESTs: LD12115.3prime  
 Contact: Stapleton, M.  
 BDGP

FEATURES  
 source  
 Location/Qualifiers  
 1..569  
 /organism="Drosophila melanogaster"  
 /db\_xref="BDGP\_EST:BDGP011306"  
 /db\_xref="taxon:7227"  
 /clone\_1db="LD Drosophila melanogaster embryo Bluescript"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="SOLR"  
 /note="Organ: embryo; Vector: Bluescript SK; Site\_1: EcoRI  
 ; Site\_2: XhoI; Constructed using Stratagene ZAP-cDNA  
 Synthesis kit. Oligo dt-primed and directionally cloned at  
 EcoRI and XhoI in Bluescript SK(+/-)"  
 BASE COUNT 127 a 161 c 143 g 138 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 194 Length: 569  
 Score: 51.00 Matches: 11  
 Percent Similarity: 53.85% Conservative: 3  
 Best Local Similarity: 42.31% Mismatches: 12  
 Query Match: 39.53% Indels: 0  
 DB: 9 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x A1388559 (1-569)  
 Oy 2 AspalaleuGUALaleuAlaaspGlnthraspalaleuGinsersGluAlaAlaVal 21  
 Db 228 GATCGCGGAATCATATAGAACACACAGACGACGAGACGAGAACCAACGCGC 169

RESULT 29  
 LOCUS A1388559/c 581 bp mRNA linear EST 19-APR-2001  
 DEFINITION GH19508 5prime GH Drosophila melanogaster head port2 Drosophila  
 melanogaster cDNA clone GH19508 5prime, mRNA sequence.  
 VERSION A1388559  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 581)  
 REFERENCE Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
 Lewis, S. and Rubin, G. M.  
 BDGP/HHMI Drosophila EST Project  
 TITLE Unpublished (2001)  
 JOURNAL Contact: Stapleton, M.  
 COMMENT BDGP

Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
hit genomic sequence AC005749  
plate: 195 row: A column: 8  
High quality sequence stop: 465.  
Location/Qualifiers

## FEATURES

1..581

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="GH19508"

/clone.lib="GH Drosophila melanogaster head POT2"

/sex="male and female"

/dev\_stage="adult"

/lab\_host="DHS - alpha"

/note="Organ: head; Vector: POT2; Site\_1: EcoRI; Site\_2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."

BASE COUNT 129 a 164 c 144 g 144 t

## ORIGIN

## Alignment Scores:

Pred. No.:	Length:	581
Score:	Matches:	11
Percent Similarity:	Conservative:	3
Best Local Similarity:	Mismatches:	12
Query Match:	Indels:	0
DB:	Gaps:	0

US-09-847-539a-6\_copy\_59\_86 (1-28) x A1108834 (1-581)

OY 2 ASPALALEUCLALALEUAlaSPGIntThASPALALEUCLInserGluGluAlaAlaVal 21

DB 227 GATCGGGCGGAAATCATATAGAACACACAGACGAGAGAGAGAGAGAAACAAAGCGCGC 168

OY 22 VALLYSALASPASnala 27

DB 167 GTACGAGAGATCGTGC 150

RESULT 30 A1108834 583 bp mRNA linear EST 19-APR-2001

LOCUS GH08128.5prtime GH Drosophila melanogaster head POT2 Drosophila

DEFINITION melanogaster cDNA clone GH08128 5prtime, mRNA sequence.

ACCESSION A1108834

VERSION A1108834.1 GI:3477369

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

AUTHORS Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 583)

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S., and Rubin, C. M.

BDGP/HIMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd. Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)

hit genomic sequence DS06378

plate: 81 row: C column: 4

High quality sequence stop: 492.

Location/Qualifiers

1..583

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="GH08128"

/clone.lib="GH Drosophila melanogaster head POT2"

/sex="male and female"

/dev\_stage="adult"

/lab\_host="DHS - alpha"

/note="Organ: head; Vector: POT2; Site\_1: EcoRI; Site\_2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."

BASE COUNT 129 a 164 c 146 g 144 t

## ORIGIN

## Alignment Scores:

Pred. No.:	Length:	583
Score:	Matches:	11
Percent Similarity:	Conservative:	3
Best Local Similarity:	Mismatches:	12
Query Match:	Indels:	0
DB:	Gaps:	0

US-09-847-539a-6\_copy\_59\_86 (1-28) x A1108834 (1-583)

OY 2 ASPALALEUCLALALEUAlaSPGIntThASPALALEUCLInserGluGluAlaAlaVal 21

DB 228 GATCGGGCGGAAATCATATAGAACACACAGACGAGAGAGAGAGAGAAACAAAGCGCGC 169

OY 22 VALLYSALASPASnala 27

DB 168 GTACGAGAGATCGTGC 151

RESULT 31 A2698116 615 bp DNA linear GSS 24-JAN-2001

LOCUS RPCI-23-224F10.TV RPCI-23 Mus musculus genomic clone RPCI-23-224F10

DEFINITION DNA sequence.

ACCESSION A2698116

VERSION A2698116.1 GI:12414882

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 615)

Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Aklnet,

B., Levins, M., Moggan, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C. M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Contact: Shaying Zhao

Other-GSS: RPCI-23-224F10.TV

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: [szhao@tigr.org](mailto:szhao@tigr.org)

Clones are derived from the mouse BAC library RPCI-23. For BAC

libraries availability, please contact Pletier de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end

page: [http://www.tigr.org/tcd/bac-ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tcd/bac-ends/mouse/bac_end_intro.html)

plate: 224 row: F column: 10

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..615

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-224F10"

/clone.lib="RPCI-23"

/sex="female"

/lab\_host="DH10B"

/note="Organ: kidney/Brain; Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into



TITLE Lewis, S. and Rubin, G. M.  
JOURNAL BDGP/HIMI Drosophila EST Project  
COMMENT Unpublished (2001)  
Contact: Stapleton, M.  
BDGP

Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
hit genomic sequence DS06378  
Plate: 21 row: 6 column: 1  
High quality sequence stop: 567.  
Location/Qualifiers

## FEATURES

source

1. 633  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="GH02173"  
/clone\_lib="GH Drosophila melanogaster head pot2"  
/sex="male and female"  
/dev\_stage="adult"  
/lab\_host="DH5 - alpha"  
/note="Organ: head; Vector: pot2; Site: 1; EcoRI: Site 2; XhoI: Sited fractionated cDNAs were directly ligated into pot2 Plasmid cDNA library."  
BASE COUNT 140 a 182 c 153 g 158 t  
ORIGIN

## Alignment Scores:

Pred. No.: 225 Length: 633  
Score: 51.00 Matches: 11  
Percent Similarity: 53.85% Conservative: 3  
Best Local Similarity: 42.31% Mismatches: 12  
Query Match: 39.53% Indels: 0  
DB: 9 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AI062828 (1-633)

QY 2 AsPAlaLeuGluAlaLeuAlaAspGlnThrAsPAlaLeuGlnSerGluGluAlaAlaVal 21

Db 227 GATCGGGCGAATCATATAGAACACACAGACGACGACGAGCAAGCAAGCGCGC 168

QY 22 VallySaLaAspAsnAla 27

Db 167 GTACGAGAGATCGTCGC 150

RESULT 34 AA391497 643 bp mRNA linear EST 19-APR-2001

LOCUS LD10172.5prtime LD Drosophila melanogaster embryo Bluescript

DEFINITION Drosophila melanogaster cDNA clone LD10172 5prtime, mRNA sequence.

ACCESSION AA391497

VERSION AA391497.1 GI:2044472

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

AUTHORS Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

TITL Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

JOURN 1 (bases 1 to 643)

COMMENT Harvey, D., Brokslein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

BDGP Lewy, S. and Rubin, G. M.

JOURN BDGP/HIMI Drosophila EST Project

COMMENT Unpublished (2001)

CONTACT: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)

Plate: 101 row: F column: 12

High quality sequence stop: 463.

Location/Qualifiers

1. 643

source

/organism="Drosophila melanogaster"  
/db\_xref="BDGP\_EST:BDCLN009443"  
/db\_xref="taxon:7227"  
/clone="LD10172"  
/clone\_lib="LD Drosophila melanogaster embryo Bluescript"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="SOLR"  
/note="Organ: embryo; Vector: Bluescript SK; Site: 1; EcoRI : Site 2; XhoI: Constructed using Stratagene ZAP-cDNA Synthesis kit. Oligo dT-primed and directionally cloned at EcoRI and XhoI in Bluescript SK(+/-)."

BASE COUNT 140 a 185 c 154 g 164 t  
ORIGIN

## Alignment Scores:

Pred. No.: 230 Length: 643  
Score: 51.00 Matches: 11  
Percent Similarity: 53.85% Conservative: 3  
Best Local Similarity: 42.31% Mismatches: 12  
Query Match: 39.53% Indels: 0  
DB: 9 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AA391497 (1-643)

QY 2 AsPAlaLeuGluAlaLeuAlaAspGlnThrAsPAlaLeuGlnSerGluGluAlaAlaVal 21

Db 213 GATCGGGCGAATCATATAGAACACACAGACGACGACGAGCAAGCAAGCGCGC 154

QY 22 VallySaLaAspAsnAla 27

Db 153 GTACGAGAGATCGTCGC 136

RESULT 35 BE337641

LOCUS BE337641 650 bp mRNA linear EST 14-JUL-2000

DEFINITION 894047D11.Y1 C. reinhardtii CC-1690, normalized, lambda zap II

ACCESSION BE337641

VERSION BE337641.1 GI:9210726

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii

REFERENCE Chlamydomonas reinhardtii

AUTHORS Chlamydomonas reinhardtii

TITL Chlamydomonas reinhardtii

JOURN Chlamydomonas reinhardtii

COMMENT Chlamydomonas reinhardtii

CONTACT: Elizabeth H. Harris

DCMB Box 91000

Duke University

Durham, NC 27708-1000, USA

Tel: 919 613 8164

Fax: 919 613 8177

Email: [chlamy@duke.edu](mailto:chlamy@duke.edu).

Location/Qualifiers

1. 650

/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mc+ 21gr"

/db\_xref="taxon:3055"

/clone\_lib="C. reinhardtii CC-1690, normalized, lambda zap II"

/note="Vector: pBluescript II SK-; Site: 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA



```
DB 167 GTACGAGAGATCGTCC 150
||||: ||| |||
RESULT 38
LOCUS A2699109/c 666 bp DNA linear GSS 24-JAN-2001
DEFINITION RPCI-23-224J19.TV RPCI-23 Mus musculus genomic clone RPCI-23-224J19
ACCESSION A2699109
VERSION A2699109.1 GI:12416865
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus;
1 (bases 1 to 666)
REFERENCE Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatman,S., Alinret,
AUTHORS B., Levins,M., McGinn,S., Tsengaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other GSS: RPCI-23-224J19.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.html). BAC end
Page: http://www.tigr.org/cdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 224 row: J column: 19
Seq primer: T7
Class: BAC ends.
FEATURES
Location/Qualifiers
source 1..666
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-224J19"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 203 a 163 c 150 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 241 Length: 666
Score: 51.00 Matches: 11
Percent Similarity: 68.18% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 7
Query Match: 39.53% Indels: 0
Gaps: 0
DB: 12
US-09-847-539a-6_COPY_59_86 (1-28) x A2699109 (1-666)
QY 1 serAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 20
DB 522 TCAGATCCCTCGAGAGACTTAAGATATCAGACAGATTCCCTAAAGCGACAGAGAGAG 463
QY 21 VALVAL 22
DB 462 GTAGTG 457
```

```
RESULT 39
LOCUS B1637204/c 710 bp mRNA linear EST 10-SEP-2001
DEFINITION SD19151.5prime SD Drosophila melanogaster Schneider L2 cell culture
PORT2 Drosophila melanogaster cDNA clone SD19151 5 similar to
CG7289: FBan0007289 GO:[] located on: 2L 22B8-22B8; 05/19/2001,
mRNA sequence.
ACCESSION B1637204
VERSION B1637204.1 GI:15539414
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 710)
REFERENCE Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
AUTHORS Lewis,S. and Rubin,G.M.
TITLE BDCP/HIMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDCP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST/estfruitfly.berkeley.edu
hit genomic AE003584: arm:2L [1824960,2149443]
estimated-cyto:22B4-22D2: 05/19/2001
Plate: SD.191 row: E column: 3
High quality sequence stop: 578.
FEATURES
Location/Qualifiers
source 1..710
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD19151"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture PORT2"
/lab_host="DH5-alpha"
/note="Vector: PORT2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into PORT2.
Plasmid cDNA library."
BASE COUNT 165 a 203 c 170 g 171 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 264 Length: 710
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
Gaps: 0
DB: 10
US-09-847-539a-6_COPY_59_86 (1-28) x B1637204 (1-710)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
DB 228 GATCGCGCGAATCATATTAAGACACACAGACGACGAGACGAGACGAGACGAGCGCG 169
QY 22 VALYVALAspAspAla 27
DB 168 GTACGAGAGATCGTCC 151
RESULT 40
LOCUS BF494445/c 797 bp mRNA linear EST 19-APR-2001
DEFINITION AT03046.5prime AT Drosophila melanogaster adult testes PORT7
Drosophila melanogaster cDNA clone AT03046 5 similar to CG7289:
FBan0007289 located on: 2L 22B8-22B8; 04/07/2001, mRNA sequence.
ACCESSION BF494445
VERSION BF494445.2 GI:13684941
KEYWORDS EST.
SOURCE fruit fly.
```



ORGANISM *Drosophila melanogaster*

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

1 (bases 1 to 797)

Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman

, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guadin, H., Harris, N., Li, P., Liao, G., Miranda, A., Mistra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragass, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celisner, S., and Rubin, G.M.

BDGP/HMI AT *Drosophila* EST Project

Unpublished (2000)

## TITLE

On Dec 6, 2000 this sequence version replaced gi:11577746.

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)

hit genomic AB003584; arm:2L [1824960,2149443]

estimated-cyto:22B4-22D2: 04/07/2001

Plate: AT.30 row: D Column: 10

High quality sequence stop: 699.

Location/Qualifiers

## FEATURES

source

1..797

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="AT03046"

/sex="male"

/dev\_stage="0-3 day old Ore-R males"

/ab\_pos="Plates AT.10-AT.120: DHS-alpha. Plates

AT.121-AT.319: DHS-alpha Tona"

/note="Organ: ADULT testes; Vector: pOTB7; Site:1: EcorI;

Site\_2: XhoI; The mRNA for the testis library was made

from testes and seminal vesicles hand dissected from 0-3

day old Ore-R males. RNA kindly provided by the lab of

Margaret Fuller. Sized fractionated cDNAs were directly

ligated into pOTB7. Plasmid cDNA library."

## BASE COUNT

178 a 234 c 195 g 190 t

## ORIGIN

## Alignment Scores:

Pred. No.: 310

Score: 51.00

Percent Similarity: 53.85%

Best Local Similarity: 42.31%

Query Match: 39.53%

DB: 10

Length: 797  
Matches: 11  
Conservative: 3  
Mismatch: 12  
Indels: 0  
Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BR494445 (1-797)

QY 2 ASPALALEUGLUALALEUALASPGLNTHRASPALALEUGINSERGIUGLUALALEUAL 21

DB 203 GATGGGGGATCATATAGAACACACAGACGACAGAGAGAGAGAGAGAGAGAGCGCG 144

QY 22 VALLYSALAASPASALA 27

DB 143 GTACGAGAGATCGTCC 126

Search completed: October 13, 2002, 06:10:16  
Job time : 333.61 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:46:02 : Search time 12.1283 Seconds  
(without alignments)  
399.383 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALFALADQTDALQSEEAAYVKADNAA 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP TREMBL\_19:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhc:\*  
9: sp.organelle:\*  
10: sp.plant:\*  
11: sp.podent:\*  
12: sp.virus:\*  
13: sp.vertibrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.dactariap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129	100.0	156	2	Q9S6G3 streptococc
2	129	100.0	156	2	Q9S6G4 streptococc
3	129	100.0	217	16	Q9X5C4 streptococc
4	129	100.0	239	2	Q9X5C5 streptococc
5	129	100.0	268	2	Q9X5C6 streptococc
6	57	44.2	377	5	076220 tetrahymena
7	57	44.2	412	16	025885 helicobacte
8	57	44.2	412	16	092507 helicobacte
9	57	44.2	850	2	Q935B1 salmonella
10	51.5	39.5	999	2	Q931X0 streptomyce
11	51	39.5	203	16	092PR3 listeria in
12	51	39.5	276	2	086553 streptomyce
13	51	39.5	486	17	Q9HNV3 halobacteri
14	51	39.5	4589	5	076506 tetrahymena
15	50	38.8	1968	5	020439 caenorhabdi
16	49.5	38.4	396	2	087677 amycolatops

17	49	38.0	207	2	Q9JP85 streptomyce
18	49	38.0	251	16	Q92P10 rhizobium m
19	49	38.0	445	10	Q9SVK1 arabidopsis
20	49	38.0	684	17	Q9HPR2 halobacteri
21	48	37.2	207	9	Q94MN6 bacterioph
22	48	37.2	1160	10	Q9PFF8 arabidopsis
23	47	36.4	254	16	084403 chlamydia t
24	47	36.4	254	16	Q9PJ29 chlamydia m
25	47	36.4	387	2	Q9KYW4 streptomyce
26	47	36.4	1617	5	Q95TX4 leishmania
27	46.5	36.0	536	17	Q9HPA2 halobacteri
28	46.5	36.0	541	17	Q979W1 thermoplas
29	46.5	36.0	749	2	Q9KZ05 streptomyce
30	46.5	36.0	784	2	086858 streptomyce
31	46	35.7	106	16	Q9K0Z7 neisseria m
32	46	35.7	208	2	082964 burkholderi
33	46	35.7	220	2	Q9K4B4 streptomyce
34	46	35.7	237	2	Q9AD77 streptomyce
35	46	35.7	250	2	Q9ZAK1 streptomyce
36	46	35.7	299	10	Q9FCP2 arabidopsis
37	46	35.7	303	10	Q9LUF2 arabidopsis
38	46	35.7	341	12	Q9IGV4 sagiyama v1
39	46	35.7	345	16	Q9RQV2 caulobacter
40	46	35.7	351	12	Q9E338 alphavirus
41	46	35.7	353	16	Q9WY02 thermotoga
42	46	35.7	367	2	Q9L034 streptomyce
43	46	35.7	526	16	Q91491 pseudomonas
44	46	35.7	539	16	Q9KFB8 bacillus ha
45	46	35.7	732	17	Q9HMP9 halobacteri

## ALIGNMENTS

RESULT 1  
ID Q9S6G3 PRELIMINARY; PRT; 156 AA.  
AC Q9S6G3:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DT GRAB (FRAGMENT).  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID-1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AP1;  
RX MEDLINE-99269061; PubMed-10336419;  
RA Rasmussen M., Muller H.P., Bjorck L.;  
RT Protein GRAB of streptococcus pyogenes regulates proteolysis at the  
RT bacterial surface by binding alpha2-macroglobulin.";  
RL J. Biol. Chem. 274:15336-15344(1999).  
DR EMBL: AF124400; AAD26339.1; -;  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
FT NON\_TER  
FT TER  
SQ SEQUENCE 156 AA; 16587 MW; 19FAA98D0599D866 CRC64;  
Query Match 100.0%; Score 129; DB 2; Length 156;  
Best Local Similarity 100.0%; Pred. No. 6.5e-10;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SDALFALADQTDALQSEEAAYVKADNAA 28  
Db 59 SDALFALADQTDALQSEEAAYVKADNAA 86  
RESULT 2  
ID Q9S6G4 PRELIMINARY; PRT; 156 AA.

```

AC 09S6G4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KT13;
RX MEDLINE-99269061; PubMed-10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
RT bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124401; AAD26340.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 156
SQ SEQUENCE 156 AA; 16575 MW; 0C20967E2FB0D866 CRC64;

Query Match . 100.0%; Score 129; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 6.5e-10;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
Db 59 SDALEALADQTDALQSEEAAYVKADNAA 86

RESULT 3
O9X5C4 PRELIMINARY; PRT; 217 AA.
ID O9X5C4;
AC O9X5C4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GRAB PRECURSOR (PROTEIN GRAB) (PROTEIN G-RELATED ALPHA 2M-BINDING
DE PROTEIN).
GN GRAB OR SPY1357.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC700294;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of Streptococcus pyogenes regulates proteolysis at the
RT bacterial surface by binding alpha2-macroglobulin.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE-21192684; PubMed-11296296;
RA Perrelet J.-J., Meschan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.;
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AF124399; AAD26338.1; -.
DR EMBL; AE006573; AKK34185.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PRINTS: PR00015; GP0SANCHOR.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome; signal; Transmembrane.
FT SIGNAL 1
FT SIGNAL 33
FT CHAIN 34 217 GRAB.
SQ

```

```

SQ SEQUENCE 217 AA; 22836 MW; 79AAB8CAF5F3FA06 CRC64;

Query Match . 100.0%; Score 129; DB 16; Length 217;
Best Local Similarity 100.0%; Pred. No. 9.4e-10;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
Db 92 SDALEALADQTDALQSEEAAYVKADNAA 119

RESULT 4
O9X5C5 PRELIMINARY; PRT; 239 AA.
ID O9X5C5;
AC O9X5C5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KT19;
RX MEDLINE-99269061; PubMed-10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
RT bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124402; AAD26341.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 239
SQ SEQUENCE 239 AA; 24846 MW; 61AC4F6F863AF0F5 CRC64;

Query Match . 100.0%; Score 129; DB 2; Length 239;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
Db 87 SDALEALADQTDALQSEEAAYVKADNAA 114

RESULT 5
O9X5C6 PRELIMINARY; PRT; 268 AA.
ID O9X5C6;
AC O9X5C6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AP49;
RX MEDLINE-99269061; PubMed-10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
RT bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124403; AAD26342.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 268
SQ SEQUENCE 268 AA; 27744 MW; 1C59239260CDC7E7 CRC64;

```

```
Query Match      100.0%; Score 129; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 1.2e-09;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SDALADQTDALQSEEAAYVKADNNA 28
    |||||
Db 171 SDALADQTDALQSEEAAYVKADNNA 198

RESULT 6
ID 076220 PRELIMINARY; PRT: 377 AA.
AC 076220;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GRANULE LATTICE PROTEIN 5.
GN GRL5.
OS Tetrathymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrathymenina; Tetrathymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B1686;
RA Verbsky J.W., Turkewitz A.P.;
RT "Proteolytic processing and Ca2+ binding activity of dense-core
RL vesicle polypeptides in Tetrathymena.";
DR Mol. Biol. Cell 0:0-0(1997).
FT CHAIN: AF031321: AAC27989.1; -.
SQ SEQUENCE 377 AA; 41509 MW; 67F52DB0295D15C CRC64;

Query Match      44.2%; Score 57; DB 5; Length 377;
Best Local Similarity 50.0%; Pred. No. 8.8;
Matches 12; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Oy 2 DALADQTDALQSEEAAYVKAD 25
    |||||
Db 133 DALADQTDALQSEEAAYVKAD 156

RESULT 7
ID 025885 PRELIMINARY; PRT: 412 AA.
AC 025885;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 47.1 KDA PROTEIN.
GN HPI327.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695 / ATCC 700392;
RC MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Loftus B., Richardson D., Dodson R., Khakhria E.F., Peterson S.,
RA McInerney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RA "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
DR EMBL: AE000634; AAD08377.1; -.

DR TIGR: HPI327; -.
DR InterPro: IPR003423; OEP.
DR Pfam: PF02321; OEP; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 412 AA; 47115 MW; 1724D849BFA15DD CRC64;

Query Match      44.2%; Score 57; DB 16; Length 412;
Best Local Similarity 52.2%; Pred. No. 9.7;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Oy 4 LEALADQTDALQSEEAAYVKADN 26
    :|||
Db 49 IQALQEQIDALSSQEKVSKMDN 71

RESULT 8
ID 092J07 PRELIMINARY; PRT: 412 AA.
AC 092J07;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PUTATIVE.
GN JHP1247.
OS Helicobacter pylori j99 (Campylobacter pylori j99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99120557; PubMed=9923682;
RX Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RL gastric pathogen Helicobacter pylori.";
DR EMBL: AE001548; AAD06820.1; -.
DR InterPro: IPR003423; OEP.
DR Pfam: PF02321; OEP; 1.
KW Complete proteome.
SQ SEQUENCE 412 AA; 47054 MW; 20E1E2904869378F CRC64;

Query Match      44.2%; Score 57; DB 16; Length 412;
Best Local Similarity 52.2%; Pred. No. 9.7;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Oy 4 LEALADQTDALQSEEAAYVKADN 26
    :|||
Db 49 IQALQEQIDALSSQEKVSKMDN 71

RESULT 9
ID 0935B1 PRELIMINARY; PRT: 850 AA.
AC 0935B1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PUTATIVE PHAGE TAIL PROTEIN.
GN HCM2.0051C.
OS Salmonella enterica subsp. enterica serovar Typh1.
OC Plasmid pHC2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=90370;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C718;
RC MEDLINE=21534947; PubMed=11677608;
RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
```

RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebalthia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davies P., Davies R.M., Dowd L., White M., Farrar J.,  
RA Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagals K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrell B.G.;  
RT "Complete genome sequence of a multiple drug resistant *Salmonella*  
RT enterica serovar Typhimurium DT104.";  
RL Nature 413:848-852(2001).  
DR EMBL: AL513384; CAD09918.1; -.  
KW Plasmid.  
SQ SEQUENCE 850 AA; 90917 MW; 21271493A07A6563 CRC64;  
Query Match 44.2%; Score 57; DB 2; Length 850;  
Best Local Similarity 59.1%; Pred. No. 21;  
Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
OY 6 ALADOTDALOSEEAAYKADNA 27  
DB 212 AASOTFAALOSEQVAAAHAADSA 233  
RESULT 10  
O93IX0 PRELIMINARY; PRT; 999 AA.  
AC O93IX0;  
DT 01-DEC-2001 (TREMblrel. 19, Created)  
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE PUTATIVE EXONUCLEASE.  
GN SCBAC36F5.11C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Saunderson D., Harris D.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Cardeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kleiser H.M., Denapalte D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL: AL592292; CAC42847.1; -.  
KW Exonuclease.  
SQ SEQUENCE 999 AA; 108134 MW; 275A16F51A85C22 CRC64;  
Query Match 39.9%; Score 51.5; DB 2; Length 999;  
Best Local Similarity 42.5%; Pred. No. 14e-02;  
Matches 17; Conservative 3; Mismatches 7; Indels 13; Gaps 2;  
OY 1 SDALALADOT-----DALOSEEAAY---VKADNA 27  
DB 716 ADAALALDDTAHRELQRLDAMQSDAAVRAVLAEDTA 755  
RESULT 11  
O92FR3 PRELIMINARY; PRT; 203 AA.  
AC O92FR3;  
DT 01-DEC-2001 (TREMblrel. 19, Created)

DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE LINO040 PROTEIN.  
GN LINO040.  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CLIP 11262 / SEROVAR 6A;  
RX PubMed=11679669;  
RA Glaser P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetoui F., Couve E., de Darvar A., Denoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussanget O.,  
RA Eutlian K.-D., Fehil H., Gomez-Lopez N., Hain T., Haut J., Jackson D.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurupkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordisk G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;  
RT "Comparative genomics of *Listeria species*.";  
RL Science 294:849-852(2001).  
DR EMBL: AL596163; CAC95273.1; -.  
DE Listeria; LINO0040;  
KW Complete proteome.  
SQ SEQUENCE 203 AA; 22611 MW; 0F53A6AC1FC1B0B CRC64;  
Query Match 39.5%; Score 51; DB 16; Length 203;  
Best Local Similarity 38.5%; Pred. No. 29;  
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;  
OY 1 SDALALADOTDALOSEEAAYKADN 26  
DB 109 AETRELADETDKLDKEDTGVEKEN 134  
RESULT 12  
O86553 PRELIMINARY; PRT; 276 AA.  
AC O86553;  
DT 01-NOV-1998 (TREMblrel. 08, Created)  
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE PUTATIVE DEHYDROGENASE.  
GN SCIF2.16C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kleiser H.M., Denapalte D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL: AL592292; CAC42847.1; -.  
KW Exonuclease.  
SQ SEQUENCE 999 AA; 108134 MW; 275A16F51A85C22 CRC64;  
Query Match 39.9%; Score 51.5; DB 2; Length 999;  
Best Local Similarity 42.5%; Pred. No. 14e-02;  
Matches 17; Conservative 3; Mismatches 7; Indels 13; Gaps 2;  
OY 1 SDALALADOT-----DALOSEEAAY---VKADNA 27  
DB 716 ADAALALDDTAHRELQRLDAMQSDAAVRAVLAEDTA 755  
RESULT 11  
O92FR3 PRELIMINARY; PRT; 203 AA.  
AC O92FR3;  
DT 01-DEC-2001 (TREMblrel. 19, Created)

DR EMBL: AL031350; CAA20507.1; -  
 DR HSSP: 064105; 10AA.  
 DR InterPro: IPR002198; ADH\_short.  
 DR Pfam: PF00106; adh\_short.1.  
 DR PRINTS: PR00080; SDRFAMILY.  
 DR PROSITE: PS00061; ADH\_SHORT; UNKNOWN\_1.  
 DR Oxidoreductase.  
 KM SEQUENCE 276 AA; 29010 MW; 30F0038B70D63C7C CRC64;

Query Match 39.5%; Score 51; DB 2; Length 276;  
 Best Local Similarity 48.0%; Pred. No. 40;  
 Matches 12; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

OY 1 SDALADOTDALOSEAAVYKAD 25  
 Db 40 ADALHADLDEIRATHSVYDVYAD 64

## RESULT 13

O9HHV3 PRELIMINARY; PRT; 486 AA.  
 AC O9HHV3;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)  
 DE L-2,4-DIAMINOBUTYRATE DECARBOXYLASE.  
 GN BDB OR VNG6211G.  
 OS Halobacterium sp. (strain NRC-1).  
 OG Plasmid pNRC200.  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 OX NCBI\_TaxID=64091;  
 RN [1]

SEQUENCE FROM N.A.  
 RX MEDLINE-20504483; PubMed-11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shroana J.,  
 RA Swartzel S., Welt D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angelvine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlschoder M., Spudis J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,  
 RT "Genome sequence of Halobacterium species NRC-1."  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC - COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC - SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND TYRDC).  
 CC EMBL: AE005155; AAC20873.1; -  
 DR InterPro: IPR002129; Pyridoxal\_dec.  
 DR Pfam: PF00282; Pyridoxal\_dec.1.  
 DR PRINTS: PR00800; YHDCRBOXLASE.  
 KM Complete proteome; Decarboxylase; Lyase; Plasmid; Pyridoxal phosphate.  
 SQ SEQUENCE 486 AA; 51092 MW; 9CBB996078AF9A16 CRC64;

Query Match 39.5%; Score 51; DB 17; Length 486;  
 Best Local Similarity 46.2%; Pred. No. 73;  
 Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 2 DALEALADOTDALOSEAAVYKADNA 27  
 Db 355 DGLGALVDRILALADVDVAGLIRADPA 380

## RESULT 14

O76506 PRELIMINARY; PRT; 4589 AA.  
 AC O76506;  
 DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE CILIARY OUTER ARM DYNEIN BETA HEAVY CHAIN.  
 GN DYH4.

OS Tetrahymena thermophila.  
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
 OC Tetrahymenina; Tetrahymena.  
 OX NCBI\_TaxID=5911;  
 RN [1]

SEQUENCE FROM N.A.  
 RX MEDLINE-99169020; PubMed-10069817;  
 RA Lee S., Wisniewski J.C., Dentler W.L., Asai D.J.,  
 RT "Gene knockouts reveal separate functions for two cytoplasmic dyneins  
 in Tetrahymena thermophila."  
 RL Mol. Biol. Cell 10:771-784(1999).  
 RN [2]

SEQUENCE FROM N.A.  
 RA Lincoln L.M., Gibson T.M., Asai D.J., Forney J.D.,  
 RT "A gene knockout reveals that dynein beta heavy chain is required in  
 Tetrahymena thermophila."  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF072878; AAC26117.1; -  
 DR InterPro: IPR004273; Dynein\_heavy.  
 DR Pfam: PF03028; Dynein\_heavy.1.  
 SQ SEQUENCE 4589 AA; 528391 MW; 7164E57C28A12BC1 CRC64;

Query Match 39.5%; Score 51; DB 5; Length 4589;  
 Best Local Similarity 60.0%; Pred. No. 8; 6e+02;  
 Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 DALEALADOTDALOSEAAV 21  
 Db 3205 EATEALAEALPALRSREAAV 3224

## RESULT 15

O20439 PRELIMINARY; PRT; 1968 AA.  
 AC O20439; Q22545;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE MYO-1 PROTEIN.  
 GN MYO-1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]

SEQUENCE FROM N.A.  
 RP MEMURRAY A.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]

SEQUENCE FROM N.A.  
 RX MEDLINE-94150718; PubMed-7906398;  
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister M., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smalton N., Smith A., Sonnenhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P., III of C.  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans."  
 RL Nature 368:32-38(1994).  
 DR EMBL: Z68119; CAA92197.1; -  
 DR EMBL: Z68117; CAA92197.1; JOINED.  
 DR EMBL: Z68117; CAA92183.1; -  
 DR EMBL: Z68119; CAA92183.1; JOINED.  
 DR HSSP: P10587; 1BR2.  
 DR InterPro: IPR001609; myosin\_head.  
 DR InterPro: IPR004009; Myosin\_N.  
 DR InterPro: IPR002928; Myosin\_tail.  
 DR InterPro: IPR000209; Peptidase\_S8.

DR Pfam: PF00063; myosin-head; 1.  
 DR Pfam: PF02736; myosin\_N; 1.  
 DR Pfam: PF01576; myosin\_tail; 2.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR PRODOM: PD000355; myosin-head; 1.  
 DR SMART: SM00242; MSC; 1.  
 DR PROSITE: PS00136; SUBTILASE-ASP; UNKNOWN\_1.  
 DR SEQUENCE 1968 AA; 225431 MW; 5A0481693CB6267 CRC64;

Query Match  
 Best Local Similarity 38.8%; Score 50; DB 5; Length 1968;  
 Matches 9; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 SDALADQTDALQSEEAAYVK 23  
 DB 1213 SDALQSLDQIEQLQKQKRIEK 1235

RESULT 16  
 087677  
 ID 087677 PRELIMINARY; PRT; 396 AA.  
 AC 087677  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE GLYCOSYLTRANSFERASE.  
 GN BGTF.  
 OS Amycolatopsis mediterranei (Nocardia mediterranei).  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales; Amycolatopsi.  
 OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae; Amycolatopsi.  
 ON NCBI\_TaxID=33910;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM5908;  
 RX MEDLINE=99318579; PubMed=10390204;  
 RA Pelzer S., Suesmuth R., Heckmann D., Recktenwald J., Huber P.,  
 RA Jung G., Wohlleben W.;  
 RA "Identification and analysis of the balhimycin biosynthetic gene  
 RT cluster and its use for manipulating glycopeptide biosynthesis in  
 RT Amycolatopsis mediterranei DSM5908.";  
 RL Antimicrob. Agents Chemother. 43:1565-1573(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM5908;  
 RA Recktenwald J., Shauky R.M., Puk O., Pfeinlg F., Keller U.,  
 RA Wohlleben W., Pelzer S.;  
 RA "The nonribosomal biosynthesis of vancomycin-type antibiotics: A  
 RT heptapeptide backbone and eight peptide synthetase modules.";  
 RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM5908;  
 RA Pfeiffer V., Nicholson G.J., Ries J., Recktenwald J., Schefer A.B.,  
 RA Shauky R.M., Schroeder J., Wohlleben W., Pelzer S.;  
 RA "A Polyketide Synthase in Glycopeptide Biosynthesis: the Biosynthesis  
 RT of the non-ribosomal Amino Acid (S)-5,5-Dihydroxyphenylglycine.";  
 RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM5908;  
 RA Puk O., Huber P., Bischoff D., Recktenwald J., Jung G.,  
 RA Suesmuth R.D., Van Pee K.H., Wohlleben W., Pelzer S.;  
 RA "Glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908:  
 RT function of a halogenase and a haloperoxidase.";  
 RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: Y16952; CAA76551.1;  
 DR TRANSFERASE.  
 SQ SEQUENCE 396 AA; 41131 MW; 1C40BDEDDDF2B72 CRC64;

Query Match  
 Best Local Similarity 38.4%; Score 49.5; DB 2; Length 396;  
 Matches 14; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

QY 1 SDALADQTDALQSEEAAYVKAD 25  
 DB 349 SDALQSLDQIEQLQKQKRIEK 374

RESULT 17  
 092P10  
 ID 092P85 PRELIMINARY; PRT; 207 AA.  
 AC 092P85;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE TRANSCRIPTIONAL REGULATOR PDHR.  
 GN PDHR.  
 OS Streptomyces seoulensis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales; Streptomyces.  
 OC Actinomycetales; Streptomyetaceae; Streptomyetaceae; Streptomyces.  
 ON NCBI\_TaxID=73044;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=99078078; PubMed=9858775;  
 RA Youn H., Kwak J., Youn H.D., Hah Y.C., Kang S.O.;  
 RA "Lipoamide dehydrogenase from streptomyces seoulensis: biochemical and  
 RT genetic properties.";  
 RL Biochim. Biophys. Acta 1388:405-418(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Youn H., Kang S.O.;  
 RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1 SIMILARITY: BELONGS TO THE GNTR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 DR EMBL: AF047034; AAF37157.1;  
 DR InterPro: IPR001525; C5\_DNA\_meth.  
 DR InterPro: IPR000524; HTH\_Gntr.  
 DR Pfam: PF00392; gntr; 1.  
 DR SMART: SM00345; HTH\_GNTR; 1.  
 DR PROSITE: PS00095; C5\_MTHASE\_2; UNKNOWN\_1.  
 DR DNA-binding; Transcription regulation.  
 KM SEQUENCE 207 AA; 22985 MW; 478003BC4DB838A4 CRC64;

Query Match  
 Best Local Similarity 38.0%; Score 49; DB 2; Length 207;  
 Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 3 ALEALADQTDALQSEEAAYVKAD 27  
 DB 97 ALEALQSEEAALPHEAALYEAQA 121

RESULT 18  
 092P10  
 ID 092P10 PRELIMINARY; PRT; 251 AA.  
 AC 092P10;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE ABC TRANSPORTER ATP-BINDING PROTEIN.  
 GN SMC00531.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 ON NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21368234; PubMed=11474104;  
 RA Gallibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,  
 RA Barloy-Hubler F., Barnett M.J., Becker A., Bolstad P., Botte G.,  
 RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,  
 RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,  
 RA Cloux S., Godtke T., Goffeau A., Golding B., Guzy J., Gurjal M.,  
 RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,  
 RA Kahn M.L., Kaiman S., Keating D.H., Kiss E., Komp C., Lelaure V.,



RA Masny D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,  
 RA Rampeger U., Surzycki R., Thebaud P., Vandenbol M.,  
 RA Vorreiter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.,  
 RT "The composite genome of the legume symbiont *Shorhizobium meliloti*."  
 RL Science 293:668-672(2001).  
 DR EMBL: AL591788; CAC46313.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 251 AA; 27570 MW; A82EFC3007A5FD6 CRC64;

Query Match 38.0%; Score 49; DB 16; Length 251;  
 Best Local Similarity 52.4%; Pred. No. 67;  
 Matches 11; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 2 DALEALADOTDALQSEAAVY 22  
 DB 181 DALKIVADGVNLRSPRAVY 201

RESULT 19  
 O9SVK1 PRELIMINARY; PRT; 445 AA.

AC O9SVK1; 01-MAY-2000 (TREMUREL. 13, Created)  
 DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMUREL. 15, Last annotation update)  
 DE HYPOTHETICAL. 50.1 KDA PROTEIN.  
 CN F19H22.10 OR A74G38910.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OK NCBI\_TaxID=3702;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,  
 RA Mayer K.F.X., Scheller C.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Arabidopsis sequencing project;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL035679; CAB38811.1; -.  
 DR EMBL: AL161594; CAB80554.1; -.  
 KW Hypothetical protein.

SQ SEQUENCE 445 AA; 50125 MW; E70B1F0962985116 CRC64;

Query Match 38.0%; Score 49; DB 10; Length 445;  
 Best Local Similarity 43.5%; Pred. No. 1.3e+02;  
 Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 5 EALADQTDALQSEAAVVKADNA 27  
 DB 78 EALASRDEALEQRDKALSERDNA 100

RESULT 20  
 O9HPR2 PRELIMINARY; PRT; 684 AA.

AC O9HPR2; 01-MAR-2001 (TREMUREL. 16, Created)  
 DT 01-MAR-2001 (TREMUREL. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)  
 DE VNG5111C.  
 GN Halobacterium sp. (strain NRC-1).

OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 OK NCBI\_TaxID=64091;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.C., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Shroga J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welt R., Goo Y.A.,  
 RA Leitner B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenberger T.A., Peck R.F., Pohlshocher M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Bhardi H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,  
 RT "Genome sequence of Halobacterium species NRC-1."  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).

DR EMBL: AE005065; AAG19805.1; -.  
 DR HSSP: P05055; ISRO.  
 DR InterPro: IPR001667; DHH.  
 DR InterPro: IPR003029; SI.  
 DR InterPro: IPR004365; tRNA\_anti.  
 DR Pfam: PF01368; DHH; 1.  
 DR Pfam: PF00575; SI; 1.  
 DR Pfam: PF01336; tRNA\_anti; 1.  
 DR SMART: SM00316; SI; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 684 AA; 73209 MW; 830AB7964BA08713 CRC64;

Query Match 38.0%; Score 49; DB 17; Length 684;  
 Best Local Similarity 46.4%; Pred. No. 2e+02;  
 Matches 13; Conservative 4; Mismatches 9; Indels 2; Gaps 1;

OY 2 DALEALADOTDAL-QSEAAVVKADNA 27  
 DB 604 DALHRRADSYDLGAGDALRVSDDA 631

RESULT 21

ID O94MN6 PRELIMINARY; PRT; 207 AA.

AC O94MN6; 01-DEC-2001 (TREMUREL. 19, Created)  
 DT 01-DEC-2001 (TREMUREL. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)  
 DE P83.

OS Bacteriophage Mx8.  
 OC Viruses.  
 OK NCBI\_TaxID=49964;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Toudier P., Walters D., Salmi D., Magrini V., Hartzell P.L.;  
 RT "Genome Organization of temperate Myxococcus phage Mx8."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF396866; AAK94418.1; -.

SQ SEQUENCE 207 AA; 21771 MW; 45B1B5B396C448875 CRC64;

Query Match 37.2%; Score 48; DB 9; Length 207;  
 Best Local Similarity 38.2%; Pred. No. 75;  
 Matches 13; Conservative 6; Mismatches 5; Indels 10; Gaps 1;

OY 5 EALADOTD-----ALQSEAAVVKADNA 28  
 DB 48 EALAEHAKAEVAVYARGLAVAEASAAEDADNA 81

RESULT 22  
 O9FHF8 PRELIMINARY; PRT; 1160 AA.

AC O9FHF8; 01-MAR-2001 (TREMUREL. 16, Created)  
 DT 01-MAR-2001 (TREMUREL. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)  
 DE DISEASE RESISTANCE PROTEIN-LIKE.



RA Redenbach M., Kieser H.M., Denaplatte D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.\*;  
 RL MOL. Microbiol. 21:77-96(1996).  
 DR EMBL: AL355774; CAB90928.1; -;  
 SO SEQUENCE 387 AA; 40633 MW; F6DE33AB67652A57 CRC64;

Query Match 36.4%; Score 47; DB 2; Length 387;  
 Best Local Similarity 73.3%; Pred. No. 2e+02;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ALEALADQTDALQ-SEEAAYVKADNA 17  
 Db 369 ALEALADQTDALQ-SEEAAYVKADNA 383

RESULT 26

ID Q95YX4 PRELIMINARY; PRT; 1617 AA.

AC Q95YX4; PRELIMINARY; PRT; 1617 AA.  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE POSSIBLE LIVER STAGE ANTIGEN.  
 GN LMI2.85.  
 OS Leishmania major.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 ON NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,  
 RA Oliver K.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL390114; CAC59787.1; -;  
 SO SEQUENCE 1617 AA; 174798 MW; D7A5E04C928756E5 CRC64;

Query Match 36.4%; Score 47; DB 5; Length 1617;  
 Best Local Similarity 48.0%; Pred. No. 9.6e+02;  
 Matches 12; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Oy 3 ALEALADQTDALQ-SEEAAYVKADNA 27  
 Db 306 ALEALADQTDALQ-SEEAAYVKADNA 330

RESULT 27

ID Q9HPA2 PRELIMINARY; PRT; 536 AA.

AC Q9HPA2; PRELIMINARY; PRT; 536 AA.  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HTR17 TRANSDUCER.  
 GN HTR17 OR VNG1733G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 ON NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahaitas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,  
 RA Swartzel J.S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leitauer B., Keller K., Cruz R., Danon M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenberger T.A., Beck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dasarma S.;  
 RT Genome sequence of Halobacterium species NRC-1.\*;  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).

DR EMBL: AE005078; AAG19968.1; -;  
 DR HSSP: P02942; 1Q07.  
 DR InterPro: IPR004089; Chemotaxis\_transducer.  
 DR InterPro: IPR003660; HAMP.  
 DR InterPro: IPR004090; Me.Chemotaxis.  
 DR Pfam: PF000015; MCPsignal; 1.  
 DR PRINTS: PR00260; CHEMTRNSDUCR.  
 DR SMART: SM00304; HAMP; 1.  
 DR SMART: SM00283; MA; 1.  
 KW Complete proteome.  
 SO SEQUENCE 536 AA; 56815 MW; 3B33D2FD8701EE04 CRC64;

Query Match 36.0%; Score 46.5; DB 17; Length 536;  
 Best Local Similarity 50.0%; Pred. No. 3.4e+02;  
 Matches 13; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

Qy 2 DALEALADQTDALQ-SEEAAYVKADNA 27  
 Db 295 DALEALADQTDALQ-SEEAAYVKADNA 319

RESULT 28

ID Q979W1 PRELIMINARY; PRT; 541 AA.

AC Q979W1; PRELIMINARY; PRT; 541 AA.  
 DT 01-OCT-2001 (TREMblrel. 18, Created)  
 DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE PYRUVATE KINASE.  
 GN TVG1076108.  
 OS Thermoplasma volcanium.  
 OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;  
 OC Thermoplasma.  
 ON NCBI\_TaxID=50339;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GSSI / DSM 4299 / JCM 9571;  
 RX MEDLINE=20570466; PubMed=11121031;  
 RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,  
 RA Kawashima T., Yamamoto Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,  
 RA Ninozaki T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;  
 RT "Archaeal adaptation to higher temperatures revealed by genomic  
 RT sequence of Thermoplasma volcanium.\*";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).  
 DR EMBL: AF000995; BAB60191.1; -;  
 DR InterPro: IPR001697; Pyruvate\_kinase.  
 DR Pfam: PF00224; PK; 1.  
 DR Pfam: PF02887; PK\_C; 1.  
 DR PRINTS: PR01050; PYRUVTKINASE.  
 DR Prodom: PD001009; pyruvate\_kinase; 1.  
 KW Kinase; Complete proteome.  
 SO SEQUENCE 541 AA; 56810 MW; A2F5CB05C39D2FE CRC64;

Query Match 36.0%; Score 46.5; DB 17; Length 541;  
 Best Local Similarity 50.0%; Pred. No. 3.4e+02;  
 Matches 12; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Oy 1 DALEALADQTDALQ-SEEAAYVKADNA 23  
 Db 278 SDITNAIIDNADALMLSESAIGK 301

RESULT 29

ID Q9K205 PRELIMINARY; PRT; 749 AA.

AC Q9K205; PRELIMINARY; PRT; 749 AA.  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE GLUTAMATE-AMMONIA-LIGASE ADENYLYLTRANSFERASE (FRAGMENT).  
 GN GINE.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;



OY 2 DALEADOTDALQSEEAAYKADN 26  
 DB 72 DAVIDCOVDALREFELATVOAHN 96

## RESULT 33

OYK4B4 PRELIMINARY; PRT; 220 AA.

AC OYK4B4: 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL 22.9 KDA PROTEIN (FRAGMENT).

OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_Taxid=1902;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).

DR EMBL: AL359214; CAB94589.1; -.

KW Hypothetical protein.

FT NON\_TER 1  
 SQ SEQUENCE 220 AA; 22937 MW; 33BE44C192CA72FC CRC64;

Query Match 35.7%; Score 46; DB 2; Length 220;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
 Matches 12; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 5 EALADOTDALQSEEAAYKADNAA 28

DB 156 EAKDLREALORADQAKADQGA 179

## RESULT 34

OYAD77 PRELIMINARY; PRT; 237 AA.

AC OYAD77: 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE HYPOTHETICAL 25.0 KDA PROTEIN.

GN SCK13.25.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_Taxid=1902;

RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);  
 RA Seeger K.J., Harris D.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RN [2]  
 RP SEQUENCE FROM N.A.

RA STRAIN-A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RN [3]  
 RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;

RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).

DR EMBL: AL512667; CAC21634.2; -.

KW Hypothetical protein.

SQ SEQUENCE 237 AA; 24957 MW; 388BC4428FEAA224 CRC64;

Query Match 35.7%; Score 46; DB 2; Length 237;  
 Best Local Similarity 58.8%; Pred. No. 1.6e+02;  
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 ALEALADOTDALQSEEA 19

DB 150 ALQALADDTLSDAERA 166

## RESULT 35

OY2AK1 PRELIMINARY; PRT; 250 AA.

AC OY2AK1: 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE HYPOTHETICAL 26.8 KDA PROTEIN (FRAGMENT).

OS Streptomyces lividans.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_Taxid=1916;

RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-TK21;  
 RA Isteagas C., Parro V., Mellado R.P.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: Y13601; CAA73927.1; -.

KW Hypothetical protein.

FT NON\_TER 250  
 SQ SEQUENCE 250 AA; 26783 MW; B678E58C3DCA4634 CRC64;

Query Match 35.7%; Score 46; DB 2; Length 250;  
 Best Local Similarity 46.4%; Pred. No. 1.7e+02;  
 Matches 13; Conservative 4; Mismatches 9; Indels 2; Gaps 1;

OY 3 ALEALADOTDALQSE--EAAVYKADNAA 28

DB 153 AREALSEOTEASKEIRELKSRRADKAA 180

## RESULT 36

OYFGP2 PRELIMINARY; PRT; 299 AA.

AC OYFGP2: 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE 11-BETA-HYDROXYSTEROID DEHYDROGENASE-LIKE.

OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Euphorbia; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC Euphorbia; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OX NCBI\_Taxid=3702;

RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA;  
 RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
 RA Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.;"

RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 CC (SDR) FAMILY.  
 DR EMBL: AB025619; BAB09144.1; -  
 DR HSSP: P50162; 1AEL.  
 DR InterPro: IPR002198; ADH\_short.  
 DR Pfam: PF00106; adh\_short; 1.  
 DR PRINTS: PR00080; SDRFAMILY.  
 DR PROSITE: PS00061; ADH\_SHORT; UNKNOWN\_1.  
 DR Oxidoreductase.  
 KW OXIDOREDUCTASE.  
 SO SEQUENCE 299 AA; 33247 MW; 77F6AC14AD01DD9D CRC64;  
 QY Query Match 35.7%; Score 46; DB 10; Length 299;  
 Best Local Similarity 41.7%; Pred. No. 2.1e+02;  
 Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
 DB 82 DRLQVAVADRCRKLGSPPVAVVRGD 105  
 RESULT 37  
 ID 09LUF2 PRELIMINARY; PRT; 303 AA.  
 AC 09LUF2;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE 11-BETA-HYDROXYSTEROID DEHYDROGENASE-LIKE.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.  
 OX NCBI\_TaxId=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asanizu E., Kotani H.,  
 RA Tabata S.;  
 RT \*Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
 RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
 RT clones.;  
 RI DNA Res. 7:31-63(2000).  
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 CC (SDR) FAMILY.  
 CC EMBL: AB023037; BAA96982.1; -  
 DR HSSP: P50162; 1AEL.  
 DR InterPro: IPR002198; ADH\_short.  
 DR Pfam: PF00106; adh\_short; 1.  
 DR PRINTS: PR00080; SDRFAMILY.  
 DR PROSITE: PS00061; ADH\_SHORT; UNKNOWN\_1.  
 KW Oxidoreductase.  
 SO SEQUENCE 303 AA; 33689 MW; 1BA39BAE1DDAC859 CRC64;  
 QY Query Match 35.7%; Score 46; DB 10; Length 303;  
 Best Local Similarity 41.7%; Pred. No. 2.1e+02;  
 Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
 DB 82 DRLQVAVADRCRKLGSPPVAVVRGD 105  
 RESULT 38  
 ID 09IGV4 PRELIMINARY; PRT; 341 AA.  
 AC 09IGV4;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE ENVELOPE GLYCOPROTEIN EI (FRAGMENT).  
 DR Saiglyama GLYCOPROTEIN EI (FRAGMENT).  
 OS Saiglyama VITUS.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
 OC Alphavirus.  
 OX NCBI\_TaxId=59303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TAIWAN;  
 RA Wang H.W., Wang Y.M., Lin H.C., Weng M.H., Lien J.C., Shiao M.F.,  
 RA Yao C.W.;  
 RT Isolation and Characterization of Saiglyama Virus, a Member of the  
 RT Semliki Forest Virus Complex, in North Taiwan.  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AF242890; AAF91319.1; -  
 DR InterPro: IPR002548; Alpha\_E1\_glycop.  
 DR Pfam: PF01589; Alpha\_E1\_glycop; 1.  
 FT NON\_TER 1  
 SO SEQUENCE 341 AA; 36532 MW; D29H6EB91DF15207 CRC64;  
 QY Query Match 35.7%; Score 46; DB 12; Length 341;  
 Best Local Similarity 41.7%; Pred. No. 2.4e+02;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;  
 DB 8 EAYDRSDVCKHDAAYKATTA 31  
 RESULT 39  
 ID 09ROY2 PRELIMINARY; PRT; 345 AA.  
 AC 09ROY2;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE PUTATIVE MANNOsylTRANSFERASE (GLYCOSYL TRANSFERASE, GROUP 1 FAMILY  
 DE PROTEIN LPSD).  
 DE LPSD OR CCI013.  
 GN LPSD OR CCI013.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 OX NCBI\_TaxId=69394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15, AND NA1000;  
 RX MEDLINE=98292737; PubMed=9620954;  
 RA Awram P., Smit J.;  
 RT \*The Caulobacter crescentus paracrystalline S-layer protein is  
 RT secreted by an ABC transporter (type I) secretion apparatus.\*;  
 RT J. Bacteriol. 180:3062-3069(1998).  
 RL [2]  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15, AND NA1000;  
 RA Awram P., Smit J.R.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15, AND NA1000;  
 RA Awram P.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Elesen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Ueberback T., Tian K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT \*Complete genome sequence of Caulobacter crescentus.\*;  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 RL EMBL: AF062345; AAF03166.1; -  
 DR EMBL: AE005779; AAK22997.1; -

DR TIGR; CC1013; -  
 DR InterPro: IPR001296; Glycos\_transf\_1.  
 DR Pfam: PF00534; Glycos\_transf\_1; 1  
 KM Transferase; Glycosyltransferase; Complete proteome.  
 SQ SEQUENCE 345 AA; 39105 MW; 5E3B5E45D4280449 CRC64;

Query Match 35.7%; Score 46; DB 16; Length 345;  
 Best Local Similarity 43.5%; Pred. No. 2.4e+02;  
 Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 6 ALADQTDALQSEEAAYVKKADNAA 28  
 |||:| | | :| | | :  
 Db 303 ALAQFDRLHADKKAATAKMGTA 325

## RESULT 40

O9E338 PRELIMINARY; PRT; 351 AA.  
 AC O9E338;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE POLYPROTEIN.  
 OS Alphavirus M1.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
 OC Alphavirus.  
 OX NCBI\_TaxID:97469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Liang G., Zhao W., Zhou G., He H., Fu S., Li L., Jin Q., Fang M.,  
 RA Hou Y.;  
 RT "Cloning and primary analysis of 3' end of two alphaviruses from  
 RT Hainan Island";  
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE268025; AAG2123.1;  
 DR InterPro: IPR002548; Alpha\_E1\_glycop.  
 DR Pfam: PF01589; Alpha\_E1\_glycop; 1.  
 SQ SEQUENCE 351 AA; 37679 MW; C1691D77E1E6FDF CRC64;

Query Match 35.7%; Score 46; DB 12; Length 351;  
 Best Local Similarity 41.7%; Pred. No. 2.5e+02;

Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 5 EALADQTDALQSEEAAYVKKADNAA 28  
 |||:| | | :| | | :  
 Db 18 EAYVDRSDVCKHDAAYKKAHTAA 41

Search completed: October 13, 2002, 04:48:07  
 Job time : 13.1283 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:33:22 : Search time 4.04278 Seconds  
(without alignments)  
268.169 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALEALADQTDALQSEEAAYKADNAA 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57.5	44.6	309	1	SANT_PLAFN
2	51	39.5	674	1	YD64_MYCPN
3	50	38.8	617	1	ASMA_ECOLI
4	50	38.8	1947	1	MYSC_CABEL
5	47.5	36.8	421	1	EXOF_RHIME
6	47	36.4	139	1	PAND_MYCTU
7	47	36.4	192	1	NIFE_FRAAL
8	47	36.4	1014	1	MLUB_MYCLE
9	47	36.4	2418	1	SPCA_HUMAN
10	46.5	36.0	544	1	KPYK_THEAC
11	46	35.7	127	1	PAND_STRAM
12	46	35.7	425	1	IM44_CABEL
13	46	35.7	763	1	HTR2_HALNI
14	46	35.7	764	1	HTR2_HALSA
15	46	35.7	787	1	YC88_MYCPN
16	46	35.7	1000	1	Y182_STRCO
17	45	34.9	165	1	MAX_BRARE
18	45	34.9	453	1	KAR1_YEAST
19	45	34.9	474	1	KCC4_RAT
20	45	34.9	476	1	PPBH_PSEAE
21	44.5	34.5	516	1	PSP3_CHLRE
22	44.5	34.5	778	1	HTR6_HALNI
23	44.5	34.5	778	1	HTR6_HALSA
24	44	34.1	251	1	ABCX_ANTSP
25	44	34.1	618	1	CHEA_LISIN
26	44	34.1	618	1	CHEA_LISMO
27	44	34.1	1938	1	MYSD_CABEL
28	43.5	33.7	388	1	MRP4_STRPY
29	43.5	33.7	420	1	YTF3_RHLIT
30	43.5	33.7	471	1	PPB_ECOLI
31	43.5	33.7	1230	1	SMC3_YEAST
32	43	33.3	248	1	RL4_HALHA
33	43	33.3	372	1	TOLA_HAEIN

34	43	33.3	587	1	UL25_HSVB	P28928 equine hepr
35	43	33.3	621	1	YQ01_CABEL	O09293 caenorhabd
36	43	33.3	627	1	DNM_RICPR	O92049 ricetelsia
37	43	33.3	697	1	MPEI_LYCES	P93203 lycopersico
38	43	33.3	759	1	METE_MYCTU	O06584 mycobacteri
39	43	33.3	794	1	YB52_MYCPN	P75034 mycoplasma
40	43	33.3	964	1	MYLS_MYCTU	O53784 mycobacteri
41	43	33.3	1966	1	MYSB_CABEL	P02566 caenorhabd
42	43	33.3	2230	1	GOG4_HUMAN	O13439 homo sapien
43	43	33.3	2415	1	SPCA_HUMAN	P08032 mus musculu
44	43	33.3	6669	1	NEBU_MOUSE	P20929 homo sapien
45	42.5	32.9	496	1	NIFE_RHISN	P55673 rhizobium s

## ALIGNMENTS

## RESULT 1

SANT\_PLAFN STANDARD: PRT: 309 AA.

AC P04928; 13-AUG-1987 (Rel. 05, Created)

DT 13-AUG-1987 (Rel. 05, last sequence update)

DT 01-JUN-1994 (Rel. 29, last annotation update)

DE S-antigen protein precursor.

OS Plasmodium falciparum (isolate NF7 / Ghana).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=5842;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=85176931; PubMed=3886159;

RA Common A.F., Saint R.B., Coppel R.L., Brown G.V., Anders R.F.,

RT Kamp D.J.;

KT "Conserved sequences flank variable tandem repeats in two S-antigen

genes of Plasmodium falciparum.";

RL Cell 40:775-783(1985).

CC -! FUNCTION: S ANTIGENS ARE SOLUBLE HEAT-STABLE PROTEINS PRESENT

CC IN THE SERA OF SOME INFECTED INDIVIDUALS.

CC -! SUBCELLULAR LOCATION: PARASITOPHOROUS VACUOLE.

CC -! POLYMORPHISM: DIVERSITY IN S-ANTIGEN IS MAINLY DUE TO POLYMORPHISM

CC IN THE REPEATITIVE REGIONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: M10130; AAA29758.1; -.

DR PIR: B22011; YAZON7.

KW Malaria; Repeat; Signal.

FT SIGNAL 1 23

FT CHAIN 24 309

FT DOMAIN 24 56

FT DOMAIN 58 86

FT DOMAIN 97 256

FT DOMAIN 257 286

FT REPEAT 257 271

FT REPEAT 272 286

FT SEQUENCE 309 AA; 33694 MW; 63574C51D62024FC CRC64;

Query Match 44.6%: Score 57.5; DB 1; Length 309;

Best Local Similarity 50.0%: Pred. No. 0.72;

Matches 14; Conservative 6; Mismatches 7; Indels 1; Gaps 1;

OY 1 SDALEAL-ADQTDALQSEEAAYKADNAA 27

DB 97 SDALEALSKDEALAKSDEAKRSDEA 124

## RESULT 2

YD64\_MYCPN STANDARD: PRT: 677 AA.  
 AC P75417:  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MPN364 (H91\_orf677).  
 GN MPN364 OR MP472.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
 OC Mycoplasmales; Mycoplasma.  
 OX NCBI\_TaxID=2104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29342 / M129;  
 RX MEDLINE=97105885; PubMed=8948633;  
 RA Himmelfeich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,  
 RA Herrmann R.;  
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 pneumoniae.";  
 RT Nucleic Acids Res. 24:4420-4449(1996).  
 RL Nucleic Acids Res. 24:4420-4449(1996).  
 CC -1- SIMILARITY: BELONGS TO THE MG185 / MG260 FAMILY.  
 CC -----  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AE00047; AAB96120.1; -  
 DR Hypothetical protein: Complete proteome.  
 KM SEQUENCE 677 AA: 75591 MW: 8785868B0679F04D CRC64:  
 SQ

Query Match 39.5%: Score 51; DB 1; Length 677;  
 Best Local Similarity 50.0%: Pred. No. 13;  
 Matches 11: Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 3 ALEALADOTDALQSEEAAYVKA 24  
 DB 486 ATEEVLQESNTLQSEEAAYVLA 507

## RESULT 3

ASMA\_ECOLI STANDARD: PRT: 617 AA.  
 AC P28249: P76390;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein asma precursor.  
 GN ASMA OR B2064.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE OF 1-379 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=96065706; PubMed=7476172;  
 RA Misra R., Miao Y.;  
 RT "Molecular analysis of asma, a locus identified as the suppressor of  
 RT OmpF assembly mutants of Escherichia coli K-12.";  
 RL Mol. Microbiol. 16:779-788(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251358; PubMed=9097040;  
 RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,  
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,  
 RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,  
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horikuchi T.;  
 RA "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 40.1-50.0 min region on the linkage map.";  
 RL DNA Res. 3:379-392(1996).  
 RN [4]

RP SEQUENCE OF 1-200 FROM N.A.  
 RX MEDLINE=92380941; PubMed=1324907;  
 RA Wang L., Weiss B.;  
 RT "dcd (dCTP deaminase) gene of Escherichia coli: mapping, cloning,  
 RT sequencing, and identification as a locus of suppressors of lethal  
 RT dut (dUTPase) mutations.";  
 RL J. Bacteriol. 174:5647-5653(1992).  
 CC -1- FUNCTION: INVOLVED IN THE INHIBITION OF ASSEMBLY OF MUTANT OMP  
 CC PROTEINS. IN GENERAL, COULD BE INVOLVED IN THE ASSEMBLY OF OUTER  
 CC MEMBRANE PROTEINS.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic (potential).  
 CC -----  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U11035; AAD14778.1; -  
 DR EMBL: AE000266; AAC75125.1; -  
 DR EMBL: D90844; BAA15917.1; -  
 DR EMBL: D90845; BAA15922.1; -  
 DR EMBL: M90069; AAA23670.1; -  
 DR PIR: B42940; B42940.  
 DR Ecogene; EGI1361; asma.  
 KW Periplasmic; Signal; Complete proteome.  
 FT SIGNAL 1 22  
 FT CHAIN 23 617 PROTEIN ASMA.  
 FT CONFLICT 354 379 NEFYRCCHANDQVGFACNYSYTAGQT (IN REF. 1).  
 FT FT  
 SQ SEQUENCE 617 AA: 68962 MW: 23AD060D864BC680 CRC64;

Query Match 38.8%: Score 50; DB 1; Length 617;  
 Best Local Similarity 45.8%: Pred. No. 16;  
 Matches 11: Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 3 ALEALADOTDALQSEEAAYVKA 26  
 DB 109 AVIQULPQTEAVRSDDAPVAPRDN 132

## RESULT 4

MYSC\_CAEEL STANDARD: PRT: 1947 AA.  
 ID MYSC\_CAEEL  
 AC P12845:  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin heavy chain C (MHC C).  
 GN Myo-2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA MEDLINE=89178677; PubMed=2926820;  
 RX Ddbb N.J., Maruyama I.N., Krause M., Karn J.;  
 RT "Sequence analysis of the complete *Caenorhabditis elegans* myosin  
 heavy chain gene family";  
 RL J. Mol. Biol. 205:603-613(1989).  
 RN [2]  
 RP SEQUENCE OF 1-21; 112-371 AND 1501-1772 FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA MEDLINE=85201409; PubMed=388374;  
 RX Karn J., Ddbb N.J., Miller D.M.;  
 RT "Cloning nematode myosin genes";  
 RL Cell Muscle Motil. 6:185-237(1985).  
 CC -1- FUNCTION: MUSCLE CONTRACTION.  
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -1- ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.  
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
 MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE  
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
 SUBFRAGMENT (S2).  
 CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN  
 C.ELEGANS.  
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X08066; CAA30855.1; -;  
 DR EMBL: M37233; AAA28121.1; -;  
 DR EMBL: M37235; AAA28122.1; -;  
 DR EMBL: M37236; AAA28123.1; -;  
 DR PIR: S05697; S05697.  
 DR HSSP: P08799; LMMN.  
 DR InterPro: IPR004009; Myosin\_N.  
 DR InterPro: IPR002928; Myosin\_tail.  
 DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00063; myosin\_head.1.  
 DR Pfam: PF02736; myosin\_N.1.  
 DR Pfam: PF01576; Myosin\_tail.1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head.1.  
 DR SMART: SM00442; Msc.1.  
 KW Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;  
 ATP-binding; Methylation; Alkylation; Multigene family.  
 FT DOMAIN 1 855 MYOSIN HEAD-LIKE  
 FT DOMAIN 856 1947 RODLIKE TAIL (S2 AND LMM DOMAINS).  
 FT DOMAIN 856 1947 COILED COIL (POTENTIAL).  
 FT NP\_BIND 174 181 ATP (BY SIMILARITY).  
 FT DOMAIN 668 690 ACTIN-BINDING.  
 FT DOMAIN 773 787 ACTIN-BINDING.  
 FT MOD\_RES 125 125 METHYLATION (TRI-) (POTENTIAL).  
 FT MOD\_RES 708 708 ALKYLATION (SH-2).  
 FT MOD\_RES 718 718 ALKYLATION (SH-2).  
 FT CONFLICT 132 132 E -> D (IN REF. 2).  
 FT CONFLICT 137 137 M -> I (IN REF. 2).  
 SQ SEQUENCE 1947 AA; 223009 MW; 6D5A595BD013627 CRC64;

Query Match 38.8%; Score 50; DB 1; Length 1947;  
 Best Local Similarity 39.1%; Pred. No. 53;  
 Matches 9; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 SDALADQTDALQSEAAVYK 23  
 DB 1213 SDALQELSDQELQKQKGRKEK 1235  
 RESULT 5  
 EXOF\_RHIME STANDARD; PRT; 421 AA.  
 AC 002728;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Exopolysaccharide production protein exof precursor.  
 GN EXOF OR RB1068 OR SMB20945.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC plasmid pSymB (megaplasmid 2).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RCR2011 / SU47;  
 RX MEDLINE=93177026; PubMed=8433670;  
 RA Mueller P., Keller M., Weng W.M., Quandt J., Arnold W., Puehler A.;  
 RT "Genetic analysis of the Rhizobium meliloti exoYF operon: ExoY is  
 RT homologous to sugar transferases and ExoQ represents a transmembrane  
 RT protein";  
 RL Mol. Plant Microbe Interact. 6:55-65(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1021;  
 RX MEDLINE=21396508; PubMed=11481431;  
 RA Finan T.M., Weidner S., Wong K., Bunthamer J., Chain P.,  
 RA Vorholster F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouy J.,  
 RA Golding B., Puehler A.;  
 RT "The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-  
 RT fixing endosymbiont Sinorhizobium meliloti";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).  
 CC -1- FUNCTION: INVOLVED IN SUCCINOGLYCAN (EPS I) SYNTHESIS. NEEDED FOR  
 CC THE ADDITION OF THE FIRST SUGAR (GALACTOSE) TO THE ISOPRENOID  
 CC CARRIER.  
 CC -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Probable).  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L05588; AAA26265.1; -;  
 DR EMBL: AL603645; CAC49468.1; -;  
 KW Exopolysaccharide synthesis; Plasmid; Periplasmic; signal;  
 KW Complete proteome.  
 FT SIGNAL 1 31 POTENTIAL.  
 FT CHAIN 32 421 EXOPOLYSACCHARIDE PRODUCTION PROTEIN  
 FT FT EXOF.  
 SQ SEQUENCE 421 AA; 45871 MW; 55D8C5B3M469EF CRC64;  
 QY 4 LEALADQTDALQSEAAVYKADNA 28  
 DB 240 LDALADLRSLQSELEALAKKAETQA 265  
 Query Match 36.8%; Score 47.5; DB 1; Length 421;  
 Best Local Similarity 53.8%; Pred. No. 23;  
 Matches 14; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

```

RESULT 6
PAND_MYCTU          STANDARD:          PRT:          139 AA.
AC 006281:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Aspartate 1-decarboxylase precursor (EC 4.1.1.11) (Aspartate alpha-
DE decarboxylase).
GN PAND OR RV3601C OR MT3706.1 OR MTC0747B.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98293987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutherford S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT *Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.*;
RL Nature 393:537-544(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmaili M.D., Salzberg S.L.,
RA Delcher A.L., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT *Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-aspartate -> beta-alanine + CO(2).
CC -1- COFACTOR: Pyruvoyl group (by similarity).
CC -1- PATHWAY: Pantothenate biosynthesis; second branch.
CC -1- SIMILARITY: BELONGS TO THE PAND FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: 295557; CAB08943.1; -
DR EMBL: AE007170; AKK48064.1; -
DR HSSP: P31664; 1AM8.
DR TIGR: MT3706.1; -
DR TubercuList: RV3601C; -
DR InterPro: IPR003190; Asp_decarbox.
DR Pfam: PF02261; Asp_decarbox; 1.
KW Pantothenate biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen;
KW Complete proteome.
FT CHAIN 1 24
FT ASPARTATE 1-DECARBOXYLASE BETA CHAIN (BY
FT SIMILARITY).
FT MOD_RES 25 25
FT CONVERTED TO A PYRUVYL GROUP (BY
FT SIMILARITY).
SQ SEQUENCE 139 AA; 14885 MW; C5HRDCLIC996ED9C6 CRC64;
Query Match 36.4%; Score 47; DB 1; Length 139;
Best Local Similarity 40.9%; Pred. No. 8.6;

```

```

Matches 9; Conservative 3; Mismatches 10; Indels 0; Gaps 0;
QY 7 LADOTDALOSEAAVYKADNAA 28
Db 32 LMDADLLEGEQVTVYDIDNGA 53
RESULT 7
NIF2_FRAAL          STANDARD:          PRT:          192 AA.
AC P46040:
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1995 (rel. 32, Last sequence update)
DT 01-OCT-1996 (rel. 34, Last annotation update)
DE Nif2 protein.
GN NIF2.
OS Frankia alni.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Frankineae; Frankiaceae; Frankia.
OX NCBI_TaxID=1859;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CP11;
RX MEDLINE=95369734; PubMed=7642138;
RA Harriott O.T., Hosted T.J., Benson D.R.;
RT *Sequences of nifx, nifw, nifz, nifB and two ORF in the Frankia
RT nitrogen fixation gene cluster.*;
RL Gene 161:63-67(1995).
CC -1- FUNCTION: UNKNOWN.
CC -1- SIMILARITY: BELONGS TO THE NIF2 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: L29299; AAC82974.1; -
KW Nitrogen fixation.
SQ SEQUENCE 192 AA; 19002 MW; F226C72F59F50B7A CRC64;
Query Match 36.4%; Score 47; DB 1; Length 192;
Best Local Similarity 52.0%; Pred. No. 12;
Matches 13; Conservative 1; Mismatches 11; Indels 0; Gaps 0;
QY 4 LEMADOTDALOSEAAVYKADNAA 28
Db 74 LEMADDAEDATQAEPAATPTAANAA 98
RESULT 8
MMLB_MYCLE          STANDARD:          PRT:          1014 AA.
AC C06079;
DT 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Putative membrane protein mmp11.
GN MPP11 OR ML2617 OR MLC622.16C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.K., Devlin K., Dutnoy S., Feltwell T., Fraser A., Hamlin N.,

```

RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon M., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrett B.G.;  
 RT "Massive gene decay in the leprosy bacillus.";  
 RL Nature 409:1007-1011(2001).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE MPM1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z95398; CAB0803.1; -  
 DR EMBL: AL583926; CAC32149.1; ALT\_INIT.  
 DR Leproma: ML2617; -  
 DR InterPro: IPR000731; HMGR\_Patched\_57M.  
 DR PROSITE: PS50156; S5D; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 13 33 POTENTIAL.  
 FT TRANSMEM 188 208 POTENTIAL.  
 FT TRANSMEM 214 234 POTENTIAL.  
 FT TRANSMEM 235 255 POTENTIAL.  
 FT TRANSMEM 279 299 POTENTIAL.  
 FT TRANSMEM 311 331 POTENTIAL.  
 FT TRANSMEM 373 393 POTENTIAL.  
 FT TRANSMEM 409 429 POTENTIAL.  
 FT TRANSMEM 530 550 POTENTIAL.  
 FT TRANSMEM 560 580 POTENTIAL.  
 FT TRANSMEM 598 618 POTENTIAL.  
 FT TRANSMEM 649 669 POTENTIAL.  
 FT TRANSMEM 671 691 POTENTIAL.  
 SQ SEQUENCE 1014 AA; 109875 MW; 58B0D797AF56D79 CRC64;  
 Query Match 36.48; Score 47; DB 1; Length 1014;  
 Best Local Similarity 62.58; Pred. No. 69;  
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 7 LADQDTALQSEAAVY 22  
 Db 932 LADVDAIPEPTAAVY 947  
 ||| ||: ||||  
 RESULT 9  
 SPCA\_HUMAN STANDARD; PRT; 2418 AA.  
 ID SPCA\_HUMAN STANDARD; PRT; 2418 AA.  
 AC P02549; Q15514;  
 DT 21-JUL-1996 (Rel. 01, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Spectrin alpha chain, erythrocyte (Erythroid alpha-spectrin).  
 GN SPRL OR SPRA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eultheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90170949; PubMed=1689726; Scarpa A.L., Coupal E., Leto T.L.,  
 RA Sahr K.E., Laurila P., Kotula L., Winkelman J.C., Spelcher D.W., Marchesi V.T.,  
 RA Linenbach A.J., Forget B.G.;  
 RA Curtis P.J.;  
 RT "The complete cDNA and polypeptide sequences of human erythroid  
 RT alpha-spectrin.";  
 RL J. Biol. Chem. 265:4434-4443(1990).  
 RN [2]  
 RP SEQUENCE OF 7-533 FROM N.A., AND VARIANTS PRO-260; PRO-261 AND  
 RP PRO-471.

RX MEDLINE=90093318; PubMed=2794061;  
 RA Agre P., Tobe T., Scarpa A., Laughinghouse K., Marchesi S.L.,  
 RA Sahr K.E., Linenbach A.J., Marchesi V.T., Forget B.G.;  
 RT "Sequence and exon-intron organization of the DNA encoding the alpha  
 RT I domain of human spectrin. Application to the study of mutations  
 RT causing hereditary elliptocytosis.";  
 RL J. Clin. Invest. 84:1243-1252(1989).  
 RN [3]  
 RP SEQUENCE OF 7-601.  
 RX MEDLINE=84087888; PubMed=6654895;  
 RA Spelcher D.W., Davis G., Marchesi V.T.;  
 RT "Structure of human erythrocyte spectrin. II. The sequence of the  
 RT alpha-I domain.";  
 RL J. Biol. Chem. 258:14938-14947(1983).  
 RN [4]  
 RP SEQUENCE OF 7-125.  
 RX MEDLINE=84087887; PubMed=6654895;  
 RA Spelcher D.W., Davis G., Yurchenco P.D., Marchesi V.T.;  
 RT "Structure of human erythrocyte spectrin. I. Isolation of the alpha-I  
 RT domain and its cyanogen bromide peptides.";  
 RL J. Biol. Chem. 258:14931-14937(1983).  
 RN [5]  
 RP SEQUENCE OF 320-450 FROM N.A.  
 RX MEDLINE=86205962; PubMed=3458204;  
 RA Linenbach A.J., Spelcher D.W., Marchesi V.T., Forget B.G.;  
 RT "Cloning of a portion of the chromosomal gene for human erythrocyte  
 RT alpha-spectrin by using a synthetic gene fragment.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:2397-2401(1986).  
 RN [6]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE=84295638; PubMed=6472478;  
 RA Spelcher D.W., Marchesi V.T.;  
 RT "Erythrocyte spectrin is comprised of many homologous triple helical  
 RT segments.";  
 RL Nature 311:177-180(1984).  
 RN [7]  
 RP IDENTIFICATION OF PROBABLE FRAMESHIFT IN 2407-2418.  
 RA Gibson T.J.;  
 RL Unpublished observations (MAR-1995).  
 RN [8]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97001215; PubMed=8844207;  
 RA Maillet P., Alloisio N., Morie L., Delannay J.;  
 RT "Spectrin mutations in hereditary elliptocytosis and hereditary  
 RT spherocytosis.";  
 RL Hum. Mutat. 8:97-107(1996).  
 RN [9]  
 RP VARIANT SER-24.  
 RX MEDLINE=94289716; PubMed=8018926;  
 RA Parquet N., Devaux I., Boulanger L., Galand C., Boylin P.,  
 RA Lecointe M.-C., Dharmy D., Garbacz W.;  
 RT "Identification of three novel spectrin alpha I/74 mutations in  
 RT hereditary elliptocytosis: further support for a triple-stranded  
 RT folding unit model of the spectrin heterodimer contact site.";  
 RL Blood 84:303-308(1994).  
 RN [10]  
 RP VARIANTS CYS-28; HIS-28; LEU-28 AND SER-28.  
 RX MEDLINE=91358728; PubMed=1679439;  
 RA Coetzer T.L., Sahr K., Pirchal J., Blacklock H., Peterson L., Koler R.,  
 RA Doyle J., Manaster J., Palek J.;  
 RT "Four different mutations in codon 28 of alpha spectrin are  
 RT associated with structurally and functionally abnormal spectrin alpha  
 RT I/74 in hereditary elliptocytosis.";  
 RL J. Clin. Invest. 88:743-749(1991).  
 RN [11]  
 RP VARIANTS SER-28 AND ARG-48.  
 RX MEDLINE=91346849; PubMed=1878597;  
 RA Floyd P.B., Gallagher P.G., Valentini L.A., Davis M., Marchesi S.L.,  
 RA Forget B.G.;  
 RT "Heterogeneity of the molecular basis of hereditary  
 RT pyropoikilocytosis and hereditary elliptocytosis associated with  
 RT increased levels of the spectrin alpha I/74-kilodalton tryptic  
 RT peptide.";

RL Blood 78:1364-1372(1991).  
 RN [12]  
 RP VARIANT SER-45.  
 RX MEDLINE-89323468; PubMed-2568862;  
 RA Lecomte M.-C., Garbarz M., Grandchamp B., Feo C., Gautero H.,  
 Deaux I., Bourrier O., Galand C., D'Auriol L., Gilbert F.,  
 Sahr K.E., Forget B.G., Bolvin P., Dherny D.;  
 RT "Sp alpha I/78: a mutation of the alpha I spectrin domain in a white  
 RL kindred with HE and HPP phenotypes.";  
 RL Blood 74:1126-1133(1989).  
 RN [13]  
 RP VARIANT PRO-207.  
 RX MEDLINE-92176375; PubMed-1541680;  
 RA Gallagher P.G., Yse W.T., Coetzer T., Lecomte M.-C., Garbarz M.,  
 Zarkowky H.S., Baruchel A., Ballas S.K., Dherny D., Palek J.,  
 Forget B.G.;  
 RT "A common type of the spectrin alpha I 46-50a-KD peptide abnormality  
 RT in hereditary elliptocytosis and pyropoikilocytosis is associated  
 RT with a mutation distant from the proteolytic cleavage site. Evidence  
 RT for the functional importance of the triple helical model of  
 spectrin.";  
 RL J. Clin. Invest. 89:892-898(1992).  
 RN [14]  
 RP VARIANT VAL-1857.  
 RX MEDLINE-93253053; PubMed-8486776;  
 RA Wilmore R., Marechal J., Morle L., Baklouti F., Philippe N.,  
 Kastally R., Kotula L., Delaunay J., Allosio N.;  
 RT "Low expression allele alpha LEV of red cell spectrin is associated  
 RT with mutations in exon 40 (alpha V/41 polymorphism) and intron 45 and  
 RT with partial skipping of exon 46.";  
 RL J. Clin. Invest. 91:2091-2096(1993).  
 RN [15]  
 RP VARIANT BARCELONA.  
 RX MEDLINE-93373367; PubMed-8364215;  
 RA dalla Venezia N., Allosio N., Forissier A., Denoroy L., Aymerich M.,  
 Vives-Corrons J.L., Besalduch J., Besson I., Delaunay J.;  
 RT "Elliptocytosis associated with the alpha 469 His-->Pro  
 RT mutation in spectrin Barcelona (alpha I/50-46b).";  
 RL Blood 82:1661-1665(1993).  
 RN [16]  
 RP VARIANT CAGLIARI.  
 RX MEDLINE-94043025; PubMed-8226774;  
 RA Sahr K.E., Coetzer T.L., Moy L.S., Derick L.H., Chishti A.H.,  
 Jarolin P., Lorenzo F., Miraglia del Giudice E., Iolascon A.,  
 Gallanelli R.;  
 RT "Spectrin Cagliari: an Ala-->Gly substitution in helix 1 of beta  
 RT spectrin repeat 17 that severely disrupts the structure and self-  
 RT association of the erythrocyte spectrin heterodimer.";  
 RL J. Biol. Chem. 268:22656-22662(1993).  
 RN [17]  
 RP VARIANTS CULO2 AND LYON.  
 RX MEDLINE-90347052; PubMed-2384601;  
 RA Morle L., Roux A.-F., Allosio N., Pothier B., Starck J., Denoroy J.,  
 Morle F., Rudiger R.-C., Forget B.G., Delaunay J., Godet J.;  
 RT "Two elliptocytogenic alpha I/74 variants of the spectrin alpha I  
 RT domain, Spectrin Culo2 (Ggt->Gtt; alpha I 40 Gly-->Val) and  
 RT spectrin Lyon (Cpt-->Ttt; alpha I 43 Leu-->Phe).";  
 RL J. Clin. Invest. 86:548-554(1990).  
 RN [18]  
 RP VARIANT JENDOUBA.  
 RX MEDLINE-92345619; PubMed-1638030;  
 RA Allosio N., Wilmore R., Morle L., Baklouti F., Marechal J.,  
 Ducluzeau M.-T., Denoroy L., Feo C., Forget B.G., Kastally R.,  
 Delaunay J.;  
 RT "Spectrin Jendouba: an alpha II/31 spectrin variant that is  
 RT associated with elliptocytosis and carries a mutation distant from  
 RT the dimer self-association site.";  
 RL Blood 80:809-815(1992).  
 RN [19]  
 RP VARIANT TUNIS.  
 RX MEDLINE-89323436; PubMed-2568861;  
 RA Morle L., Morle F., Roux A.F., Godet J., Forget B.G., Denoroy L.,  
 Garbarz M., Dherny D., Kastally R., Delaunay J.;

RT "Spectrin Tunis (Sp alpha I/78): an elliptocytogenic variant, is due  
 RT to the CCG->TCG codon change (Arg->Trp) at position 35 of the  
 RT alpha I domain.";  
 RL Blood 74:828-832(1989).  
 RN [20]  
 RP VARIANT GENOVA.  
 RX MEDLINE-94250920; PubMed-8193371;  
 RA Perrotta S., del Giudice E.M., Allosio N., Sclarratta G., Pinto L.,  
 RA Delaunay J., Cutlilo S., Iolascon A.;  
 RT "Mild elliptocytosis associated with the alpha 34 Arg-->Trp mutation  
 RT in spectrin Genova (alpha I/74).";  
 RL Blood 83:3346-3349(1994).  
 RN [21]  
 RP VARIANT ANASTASIA.  
 RX MEDLINE-95290423; PubMed-7772539;  
 RA Perrotta S., Iolascon A., de Angelis F., Pagano L., Colonna G.,  
 RA Cutlilo S., del Giudice E.M.;  
 RT "Spectrin Anastasia (alpha I/78): a new spectrin variant (alpha 45  
 RT Arg-->Thr) with moderate elliptocytogenic potential.";  
 RL Br. J. Haematol. 89:933-936(1995).  
 CC -1- FUNCTION: SPECTRIN IS THE MAJOR CONSTITUENT OF THE CYTOSKELETAL  
 CC NETWORK UNDERLYING THE ERYTHROCYTE PLASMA MEMBRANE. IT ASSOCIATES  
 CC WITH BAND 4.1 AND ACTIN TO FORM THE CYTOSKELETAL SUPERSTRUCTURE OF  
 CC THE ERYTHROCYTE PLASMA MEMBRANE.  
 CC -1- SUBUNIT: COMPOSED OF NONHOMOLOGOUS CHAINS, ALPHA AND BETA, WHICH  
 CC AGGREGATE SIDE-TO-SIDE IN AN ANTIPARALLEL FASHION TO FORM DIMERS,  
 CC TERRAMERS, AND HIGHER POLYMERS.  
 CC -1- DISEASE: HEREDITARY ELLIPTOCYTOSIS (HE) IS A HETEROGENEOUS  
 CC DISORDER CHARACTERIZED BY VARIABLE HEMOLYTIC ANEMIA AND ELLIPTICAL  
 CC RED CELL SHAPE. SEVERAL ABNORMALITIES IN THE MEMBRANE SKELETON  
 CC HAVE BEEN IDENTIFIED IN HE, INCLUDING A NUMBER THAT HAVE BEEN

Query Match 36.4%; Score 47; DB 1; Length 2418;  
 Best Local Similarity 40.9%; Pred. No. 1.7e+02;  
 Matches 9; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Oy 2 DALEALADDTALDSEEAIVK 23  
 Db 953 DSMKALRNQANACQOQAAAPV 974  
 1::111::1:1:1111:

RESULT 10  
 KPYK\_THEAC STANDARD; PRT; 544 AA.  
 AC P32044;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pyruvate kinase (EC 2.7.1.40) (PK).  
 GN TA0896.  
 OS Thermoplasma acidophilum.  
 OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmataceae;  
 CC Thermoplasma  
 NCBI\_TaxID=2303;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-DSM 1728;  
 RC MEDLINE-20479972; PubMed-11029001;  
 RA Ruepp A., Granel W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Hewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT "The genome sequence of the thermophilic scavenger Thermoplasma  
 RT acidophilum.";  
 RL Nature 407:508-513(2000).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE-93051217; PubMed-1426985;  
 RA Potter S., Forcheggili-Gilmore L.A.;  
 RT "Purification and properties of pyruvate kinase from Thermoplasma  
 RT acidophilum.";  
 RL FEMS Microbiol. Lett. 73:235-239(1992).  
 CC -1- CATALYTIC ACTIVITY: ATP + pyruvate = ADP + phosphoenolpyruvate.  
 CC -1- COFACTOR: REQUIRES MAGNESIUM AND POTASSIUM.  
 CC -1- PATHWAY: FINAL STEP IN GLYCOLYSIS.

```

CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: BELONGS TO THE PYRUVATE KINASE FAMILY.
CC -1- CAUTION: REF.2 HAS PUBLISHED SOME PARTIAL SEQUENCE. THESE
CC SEQUENCES DO NOT ORIGINATE FROM T. ACTIOPHILUM, RATHER THEY SEEM TO
CC BE CONTAMINATED WITH HUMAN SAMPLES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AL45065; CAC12025.1; -.
CC InterPro: IPR001697; Pyruvate_kinase.
CC Pfam: PF00224; PK.1.
CC PRINTS: PR01050; PYRUVTKINASE.
CC PRODOM: PD001009; Pyruvate_kinase; 1.
CC PROSITE: PS00110; PYRUVATE_KINASE; MAGNESIUM; FALSE NEG.
CC TRANSFERASE: Kinase; Glycolysis; Magnesium; Complete proteome.
CC ACT_SITE 202 202 BY SIMILARITY.
CC METAL 204 204 MAGNESIUM (BY SIMILARITY).
CC METAL 225 225 MAGNESIUM (BY SIMILARITY).
CC METAL 226 226 MAGNESIUM (BY SIMILARITY).
CC SEQUENCE 544 AA; 59147 MW; 7BD8BCAF4706A5B2 CRC64;

Query Match 36.08; Score 46.5; DB 1; Length 544;
Best Local Similarity 50.08; Pred. No. 42;
Matches 12; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

QY 1 SDALALADQTDALO-SEEAAYK 23
Db 278 SDITNAIIDNALMLSESAIGK 301

RESULT 11
PAND_STAMM STANDARD; PRT; 127 AA.
AC 099R40;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Aspartate 1-decarboxylase precursor (EC 4.1.1.11) (Aspartate alpha-
DE decarboxylase).
GN PAND OR SA2597 OR SA2390.
OS Staphylococcus aureus (Strain M50 / ATCC 700699), and
OS Staphylococcus aureus (Strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_Taxid=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=1418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iino T.,
RA Kanemori M., Matsuura H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-U Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hiraoka H., Kihara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RA Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.
RT Lancet 357:1225-1240(2001).
CC -1- CATALYTIC ACTIVITY: L-aspartate -> beta-alanine + CO(2).
CC -1- COFACTOR: Pyruvoyl group. (By similarity).
CC -1- PATHWAY: Pantothenate biosynthesis; second branch.
CC -1- SIMILARITY: BELONGS TO THE PAND FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AP003365; BAB58759.1; ALT_INIT.
CC EMBL: AP003137; BAB43695.1; ALT_INIT.
CC InterPro: IPR003190; Asp.decarbox.
CC Pfam: PF02261; Asp.decarbox.1.
CC TRANSFERENCE: Lyase; Decarboxylase; Pyruvate; Zymogen;
CC Complete proteome.
CC CHAIN 1 24
CC FT CHAIN 25 127 ASPARTATE 1-DECARBOXYLASE BETA CHAIN (BY
CC FT MOD_RES 25 25 SIMILARITY).
CC FT MOD_RES 25 25 CONVERTED TO A PYRUVYL GROUP (BY
CC FT SEQUENCE 127 AA; 14050 MW; F3F5BF82DB7C47EA CRC64;

Query Match 35.78; Score 46; DB 1; Length 127;
Best Local Similarity 42.98; Pred. No. 11;
Matches 12; Conservative 5; Mismatches 7; Indels 4; Gaps 1;

QY 1 SDLEALADQTDALQSEEAAYKADNAA 28
Db 30 SDLEAV-----DILPNEKVAIVNNNGA 53

RESULT 12
IM44_CAEEL STANDARD; PRT; 425 AA.
AC 002161;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable import inner membrane translocase subunit TIM44,
DE mitochondrial precursor.
GN T09B4.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Langston E., Wohlmann P., Gillam B.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN PROTEIN IMPORT INTO THE MITOCHONDRION.
CC PROBABLY INVOLVED IN TRANSLLOCATION ACROSS THE INNER MEMBRANE. AS A
CC BINDING PROTEIN REQUIRED FOR DRIVING THE IMPORT OF PREPROTEINS.
CC RECRUITS MITOCHONDRIAL HSP70 TO DRIVE PROTEIN TRANSLLOCATION INTO
CC THE MATRIX USING ATP AS AN ENERGY SOURCE (BY SIMILARITY).
CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
CC LEAST 3 DIFFERENT PROTEINS (TIM17, TIM23, TIM44) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane (Potential).
CC -1- SIMILARITY: BELONGS TO THE TIM44 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U97405; AAB53011.1; -.
CC WormPeP: T09B4.9; CE13473.
CC Mitochondrion; Inner membrane; Transport; Protein transport;
CC TRANSLOCATION; Transl peptid.
CC FT CHAIN 1 425 MITOCHONDRION.
CC FT CHAIN 2 425 PROBABLE IMPORT INNER MEMBRANE
CC FT SEQUENCE 425 AA; 49398 MW; 203DFBD614E099F8 CRC64;

```

Query Match 35.7%; Score 46; DB 1; Length 425;  
 Best Local Similarity 41.7%; Pred. No. 38;  
 Matches 10; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

OY 5 EALADOTDALOSEAAVAKDNA 28  
 DB 182 EALTRKTDGDFLEKERVANDSA 205

RESULT 13  
 HTR2\_HALNI STANDARD: PRT: 763 AA.

ID HTR2\_HALNI  
 AC Q9HP81; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis protein II) (MPP-II).  
 GN HTR2 OR VNG1765G.  
 OS Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium sp. (strain NRC-1).  
 CC Halobacterium  
 CC NCBI\_Taxid=64091;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng M.V., Kennedy S.P., Maharas G.G., Bergquist B., Pan M., Shukla H.D., Lasky S.R., Baliga S.S., Thorsson V., Schroga J., Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A., Leihner B., Keller K., Cruz R., Danson M.J., Hough D.W., Maddocks D.G., Jablonki P.E., Krebs M.P., Angelino C.M., Dale H., Aisenbarger T.A., Peck R.F., Pohlschoder M., Spudis J.L., Jung K.-H., Alam M., Freltas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D., Eberhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassaime S.;  
 RA "Genome sequence of Halobacterium species NRC-1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: AE005080; AAC19989.1;  
 DR InterPro: IPR004089; Chemotaxis\_transducer.  
 DR InterPro: IPR003660; HAMF.  
 DR Pfam: PF00672; HAMF. 1.  
 DR Pfam: PF00015; MCPsignal. 1.  
 DR SMART: SM00304; HAMF. 2.  
 DR SMART: SM00283; MA. 1.  
 KW Transducer; Photoreceptor; Transmembrane; Methylation;  
 KW Complete proteome.  
 FT INIT\_MET 0  
 FT DOMAIN 1 15  
 FT TRANSMEM 16 36  
 FT DOMAIN 37 277  
 FT TRANSMEM 278 297  
 FT DOMAIN 298 763  
 FT SEQUENCE 763 AA; 78911 MW; CF7ABFF04DFE309A CRC64;

Query Match 35.7%; Score 46; DB 1; Length 763;  
 Best Local Similarity 52.4%; Pred. No. 70;  
 Matches 11; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 1 SDALEALADOTDALOSEAAV 21

DB 550 ADDAEAAADAMALDSEMAI 570

RESULT 14  
 HTR2\_HALSA STANDARD: PRT: 764 AA.

ID HTR2\_HALSA  
 AC P71410; 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis protein II) (MPP-II).  
 GN HTR2 OR HTRII.  
 OS Halobacterium salinarum.  
 CC Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 CC Halobacterium  
 CC NCBI\_Taxid=2242;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FLX15;  
 RX MEDLINE=96323203; PubMed=8710852;  
 RA Zhang W., Brooun A., Mueller M.M., Alam M.;  
 RT "The primary structures of the Archaeon Halobacterium salinarum blue light receptor sensory rhodopsin II and its transducer, a methyl-accepting protein.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:8230-8235(1996).  
 CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.  
 CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: U62676; AAC4369.1;  
 DR HSP: P02942; 1007.  
 DR InterPro: IPR004089; Chemotaxis\_transducer.  
 DR InterPro: IPR003660; HAMF.  
 DR Pfam: PF00672; HAMF. 1.  
 DR Pfam: PF00015; MCPsignal. 1.  
 DR SMART: SM00304; HAMF. 2.  
 DR SMART: SM00283; MA. 1.  
 KW Transducer; Photoreceptor; Transmembrane; Methylation;  
 KW Complete proteome.  
 FT INIT\_MET 0  
 FT DOMAIN 1 15  
 FT TRANSMEM 16 36  
 FT DOMAIN 37 277  
 FT TRANSMEM 278 298  
 FT DOMAIN 299 764  
 FT SEQUENCE 764 AA; 79187 MW; 1E0D7B4E460F588 CRC64;

Query Match 35.7%; Score 46; DB 1; Length 764;  
 Best Local Similarity 52.4%; Pred. No. 70;  
 Matches 11; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 1 SDALEALADOTDALOSEAAV 21

DB 551 ADDAEAAADAMALDSEMAI 571

RESULT 15  
 YC88\_MYCPN STANDARD: PRT: 787 AA.  
 ID YC88\_MYCPN  
 AC P75489; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)



```

DE Hypothetical lipoprotein MPN288 precursor (A65_orf870).
GN MPN288 OR MP547.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129.
RX MEDLINE-97103885; PubMed-8948633;
RA Himmelfeich R., Hilbert H., Plagens H., Pirk E., Li B.-C.,
RA Hermann R.;
RT *Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE MG185 / MG260 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000053; AAB96195.1;
KW Hypothetical protein; Lipoprotein; Membrane; Signal;
KW Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 787 HYPOTHETICAL LIPOPROTEIN MPN288.
FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 787 AA; 86889 MW; FF6237DD18D3A05 CRC64;

Query Match 35.7%; Score 46; DB 1; Length 787;
Best Local Similarity 45.5%; Pred. No. 72;
Matches 10; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 3 ALPALDQTDALQSEEAAYVKA 24
DB 599 ATREVLQSTQSDRAIVLPA 620

RESULT 16
Y182_STRCO STANDARD: PRT; 1000 AA.
AC 09FC14; 09K4B4;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein 2SC3B6.28/SC7E4.01.
GN 2SC3B6.28 OR SC7E4.01.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D., Cerdano A.M., Parkhill J., Barrel B.G.,
RA Rajandream M.A.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0182 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: AL390968; CAC01334.1;
DR EMBL: AL359214; CAB94589.1;
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 26 48 POTENTIAL.
FT TRANSMEM 70 92 POTENTIAL.
FT TRANSMEM 121 143 POTENTIAL.
FT TRANSMEM 177 199 POTENTIAL.
FT TRANSMEM 220 237 POTENTIAL.
FT TRANSMEM 267 289 POTENTIAL.
FT TRANSMEM 296 318 POTENTIAL.
SQ SEQUENCE 1000 AA; 109554 MW; C30AB349CA9E410 CRC64;

Query Match 35.7%; Score 46; DB 1; Length 1000;
Best Local Similarity 50.0%; Pred. No. 93;
Matches 12; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 5 EALADQTDALQSEEAAYVKA 28
DB 936 EAOKDLEALQRAEDAKAKDQCA 959

RESULT 17
MAX_BRAVE STANDARD: PRT; 165 AA.
AC P52161;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MAX protein.
GN MAX.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93233639; PubMed-8474440;
RA Schreiber-Agus N., Horner J., Torres R., Chiu F.-C., DePinho R.A.;
RT "Zebra fish myc family and max genes: differential expression and
RT oncogenic activity throughout vertebrate evolution."
RL Mol. Cell. Biol. 13:2765-2775(1993).
CC -1- FUNCTION: TRANSCRIPTION REGULATOR. FORMS A SPOUNCE-SPECIFIC
CC DNA-BINDING PROTEIN COMPLEX WITH MYC OR MAD WHICH RECOGNIZES
CC THE CORE SEQUENCE 5'-CAC(GA)TG-3'. THE MYC-MAX COMPLEX IS A
CC TRANSCRIPTIONAL ACTIVATOR, WHEREAS THE MAD-MAX COMPLEX IS A
CC REPRESSOR (BY SIMILARITY).
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS A HETERODIMER WITH MYC OR MAD.
CC -1- SUBCELLULAR LOCATION: Nucleus.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE KIDNEY, GILLS AND
CC UTERUS. IT IS ALSO FOUND IN THE BRAIN AND HEART.
CC -1- PTM: PHOSPHORYLATED (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L11711; AAA02483.1;
DR HSSP: P25912; 1HL0.
DR ZFIN: ZDB-GENE-990415-152; max.
DR InterPro: IPR003015; HLH_Myc.
DR InterPro: IPR001092; HLH_dim.
DR Pfam: PF00010; HLH; 1.

```



```

CC -1- PTM: THE N-TERMINAL OF CALSPERMIN IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CAMK SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M63334; AAA40865.1; -
DR EMBL: M74488; AAA40845.1; ALT_SEQ.
DR EMBL: M64757; AAA40856.1; -
DR EMBL: M64757; AAA40857.1; -
DR EMBL: J04460; AAA41867.1; -
DR EMBL: J04446; AAA40990.1; -
DR PIR: A41103; TVRNC4.
DR HSP: P00518; LPHK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_kinase.
DR Pfam: PF00069; pkinase; 1.
DR SMART: SM00220; S_TKC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase: Serine/threonine-protein kinase: ATP-binding;
KW Calmodulin-binding; Testis; Alternative splicing.
FT CHAIN 306 474
FT DOMAIN 42 296
FT NP_BIND 48 56
FT BINDING 71 71
FT ACT_SITE 160 160
FT DOMAIN 338 337
FT DOMAIN 333 399
FT DOMAIN 403 413
FT CONFLICT 372 372
FT CONFLICT 474 AA; 53133 MW; 56F71AC5644DE23 CRC64;
SQ SEQUENCE
Query Match 34.9%; Score 45; DB 1; Length 474;
Best Local Similarity 58.8%; Pred. No. 58;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
OY 2 DALEALADQDAQSE 18
Db 378 EAAKAADETRKQSEE 394

```

```

RT *Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.*;
RT Nature 406:959-964(2000).
RN (2)
RP SEQUENCE OF 27-45.
RC STRAIN-H103;
RX MEDLINE-93202452; PubMed-8454193;
RA Tan A.S.P., Morobec E.A.;
RT Isolation and characterization of two immunochemically distinct
RT alkaline phosphatases from Pseudomonas aeruginosa.*;
RT FEMS Microbiol. Lett. 106:281-286(1993).
CC -1- FUNCTION: HAS ONLY A PHOSPHOMONESTERASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O -> an
CC alcohol + phosphate.
CC -1- COFACATOR: METALLOENZYME CONTAINING TWO ZINC ATOMS AND A MAGNESIUM
CC ION.
CC -1- SUBCELLULAR LOCATION: SECRETED AND PERIPLASMIC.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED CONSTITUTIVELY.
CC -1- SIMILARITY: BELONGS TO THE ALKALINE PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE004751; AAG06684.1; -
DR InterPro: IPR001952; Alk_phosphatase.
DR Pfam: PF00245; alk_phosphatase; 1.
DR PRINTS: PR00113; ALKPHPHASE.
DR PRODOM: PD001668; Alk-phosphatase; 1.
DR SMART: SM00098; alkPec; 1.
DR PROSITE: PS00123; ALKALINE_PHOSPHATASE; 1.
KW Hydrolyase; Zinc; Magnesium; Periplasmic; Signal; Complete proteome.
FT CHAIN 1 26
FT SIGNAL 1 26
FT CONFLICT 27 476
FT CONFLICT 27 27
FT CONFLICT 39 39
FT CONFLICT 44 45
FT CONFLICT 44 45
SQ SEQUENCE 476 AA; 50408 MW; A3059B30BBA809C CRC64;
Query Match 34.9%; Score 45; DB 1; Length 476;
Best Local Similarity 43.3%; Pred. No. 59;
Matches 13; Conservative 5; Mismatches 10; Indels 2; Gaps 1;
OY 1 SDALEALADQDA--LQSEBAAYVKADNAA 28
Db 330 SKAIELKDNPNNGFPVQVEGASIDKQDHAA 359

```

```

RESULT 20
PPBH_PSEAE STANDARD; PRT; 476 AA.
AC P35483; Q9HYU7;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alkaline phosphatase H precursor (EC 3.1.3.1) (High molecular weight
DE phosphatase) (H-AP).
GN PHOA OR PA3296.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PAO1;
RX MEDLINE-20437337; PubMed-10984043;
RA Stover C.K., Pham X.-O.T., Ervin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Gardner R.L., Goltier L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;

```

```

RESULT 21
RSP3_CHLRE STANDARD; PRT; 516 AA.
ID RSP3_CHLRE
AC P12759;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Radial spoke protein 3.
GN RSP3.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=1055;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-21GR.
RX MEDLINE-89308863; PubMed-2745550;
RA Williams B.D., Velleca M.A., Curry A.M., Rosenbaum J.L.;
RT "Molecular cloning and sequence analysis of the Chlamydomonas gene
RT coding for radial spoke protein 3: flagellar mutation pf-14 is an
RT ochre allele.";

```



FT DOMAIN 48 296 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 297 317 POTENTIAL.  
 FT DOMAIN 318 778 CYTOPLASMIC (POTENTIAL).  
 SO SEQUENCE 778 AA: 82077 MW: 134C7DF0A3334CD CRC64;

Query Match 34.5%; Score 44.5; DB 1; Length 778;  
 Best Local Similarity 39.3%; Pred. No. 1.1e+02;  
 Matches 11; Conservative 4; Mismatches 12; Indels 1; Gaps 1;

QY 2 DALEALADOTDLOSE-FAAIVKADNA 28  
 DB 511 DOLEAVASETDEMSATIEEVAASADOVA 538

RESULT 24  
 ABCX\_ANTSP STANDARD; PRT; 251 AA.  
 AC 002856;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Probable ATP-dependent transporter ycf16.  
 GN YCF16.  
 OS Anthlammon sp.  
 OG Chloroplast.  
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Ceramiales; Ceramiales;  
 OC Anthlammon.  
 OX NCBI\_TaxID=2767;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LB 95.79;  
 RX MEDLINE=93021132; PubMed=1404401;  
 RA Kostrewa M., Zetsche K.;  
 RT "Large ATP synthase operon of the red alga Anthlammon sp. resembles the corresponding operon in cyanobacteria.";  
 RL J. Mol. Biol. 227:961-970(1992).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. YCF16 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X63382; CAA44985.1; -  
 DR PIR: S37635; S37635.  
 DR Mendel: 6183; ANTSP:ycf16.1.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR ProSITE: PS00211; ABC\_TRANSPORTER; 1.  
 KM ATP-binding; Transport: Chloroplast.  
 FT NP\_BIND 39 46  
 SO SEQUENCE 251 AA: 28205 MW: 3707DDI736A73419 CRC64;

Query Match 34.1%; Score 44; DB 1; Length 251;  
 Best Local Similarity 33.3%; Pred. No. 41;  
 Matches 7; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

QY 2 DALEALADOTDLOSEFAAV 22  
 DB 184 DALTKIAKOINSIKTQENSII 204

RESULT 25  
 CHEA\_LISTIN STANDARD; PRT; 618 AA.  
 ID CHEA\_LISTIN  
 AC 092DW2;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Chemotaxis protein cheA (EC 2.7.3.-).  
 GN CHEA OR LIN0700.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / Serovar 6a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,  
 RA Doman E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurel O.,  
 RA Entian K.-D., Esli H., Garcia-del Portillo F., Garrido P.,  
 RA Gauthier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kuraprak G.,  
 RA Medueno E., Maltournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;  
 RT "Comparative genomics of Listeria species.";  
 RL Science 294:849-852(2001).  
 CC -1- FUNCTION: INVOLVED IN THE TRANSMISSION OF SENSORY SIGNALS FROM THE CHEMORECEPTORS TO THE FLAGELLAR MOTORS. CHEA IS AUTOPHOSPHORYLATED; IT CAN TRANSFER ITS PHOSPHATE GROUP TO EITHER CHEB OR CHEY (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE KINASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AL596166; CAC95932.1; -  
 DR Listlist: LIN00700;  
 KM Chemotaxis; Complete proteome; Kinase; Phosphorylation;  
 FT MOD\_RES 45 45 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SO SEQUENCE 618 AA: 66668 MW: 4302DAA5BDAB458 CRC64;

Query Match 34.1%; Score 44; DB 1; Length 618;  
 Best Local Similarity 41.7%; Pred. No. 1.1e+02;  
 Matches 15; Conservative 4; Mismatches 5; Indels 12; Gaps 2;

QY 4 LEAL-ADOTD-----ALOSEFAAVKADNA 27  
 DB 119 LEALHPQETDVAEYTRAIQIEEAALKAVRA 154

RESULT 26  
 CHEA\_LISTMO STANDARD; PRT; 618 AA.  
 ID CHEA\_LISTMO  
 AC 048768;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Chemotaxis protein cheA (EC 2.7.3.-).  
 GN CHEA OR LM00692.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Listeria.  
 OX NCBI\_TaxID=1639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=12067;  
 RX MEDLINE=95102110; PubMed=7803815;



```

DR EMBL: 271261; CAA95848.1; JOINED.
DR EMBL: 271261; CAA95806.1; -.
DR EMBL: 271266; CAA95806.1; JOINED.
DR PIR: S02772; MKMKL.
DR HSSP: P08799; 1MND.
DR WormPep; R0667.10; CE06253.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head.
DR Pfam; PF02736; Myosin_N.1.
DR Pfam; PF01576; Myosin_tail.1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head.1.
DR SMART; SM00242; Myosin.1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 845 MYOSIN HEAD-LIKE.
FT DOMAIN 846 1938 ALPHA-HELICAL TAILPIECE (SHORT S2).
FT DOMAIN 1171 1938 LIGHT MEROMYOSIN (LAM).
FT NP_BIND 177 184 COILED COIL (POTENTIAL).
FT DOMAIN 660 682 ATP (BY SIMILARITY).
FT DOMAIN 764 778 ACTIN-BINDING.
FT MOD_RES 128 128 METHYLATION (TR1-).
FT MOD_RES 700 700 ALKYLATION (SH-2).
FT MOD_RES 710 710 ALKYLATION (SH-2).
FT CONFLICT 94 94 F -> E (IN REF. 2).
FT CONFLICT 98 98 A -> R (IN REF. 2).
FT CONFLICT 377 377 V -> D (IN REF. 4).
FT CONFLICT 389 390 DV -> GD (IN REF. 2).
FT CONFLICT 391 391 V -> D (IN REF. 4).
FT CONFLICT 408 408 W -> N (IN REF. 2).
FT CONFLICT 474 474 Q -> G (IN REF. 2).
FT CONFLICT 577 577 L -> F (IN REF. 4).
FT CONFLICT 681 681 I -> N (IN REF. 4).
FT CONFLICT 1373 1373 S -> D (IN REF. 2).
FT CONFLICT 1659 1659 E -> Q (IN REF. 3).
SQ SEQUENCE 1938 AA; 223255 MW; 387399C8F63ACF4 CRC64;

Query Match 34.1%; Score 44; DB 1; Length 1938;
Best Local Similarity 39.1%; Pred. No. 3.5e+02;
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 1 SDALADQTDALQSEPAAYK 23
DB 1204 NDQISALTNTLDALQSKAKIEK 1226

RESULT 28
MRP4_STRPY STANDARD; PRT; 388 AA.
ID MRP4_STRPY
AC P30141;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Fibrinogen- and Ig-binding protein precursor (MRP protein).
GN MRP4.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN NCBI_TaxID=1314;
RP SEQUENCE FROM N.A.
RC STRAIN=AP4 / SEROTYPE M4;
RX MEDLINE=92269579; PubMed=1588817;
RA Stenberg L., O'Toole P., Lindahl G.;
RT "Many group A streptococcal strains express two different
RT immunoglobulin-binding proteins, encoded by closely linked genes:
RT characterization of the proteins expressed by four strains of
RT different M-type.";
RL Mol. Microbiol. 6:1185-1194(1992).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AP4 / SEROTYPE M4;
RX MEDLINE=92409576; PubMed=1528877;
RA O'Toole P., Stenberg L., Rissler M., Lindahl G.;
RT "Two major classes in the M protein family in group A streptococci.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8661-8665(1992).
CC -1- FUNCTION: BINDS IGG MOLECULES OF THE IGL, IGH AND IGA SUBCLASSES,
CC AND ALSO BINDS FIBRINOGEN.
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein. Cell wall.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M87831; AAA26930.1; -.
DR PIR: A46173; A46173.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor.1.
DR PRINTS: PR00015; GPOSANCHOR.
DR PROSITE: PS00343; GRAM_POS_ANCHORING.1.
KW IgG-binding protein; Repeat; Transmembrane; Cell wall; Signal.
FT SIGNAL 1 41
FT CHAIN 42 388 FIBRINOGEN- AND IG-BINDING PROTEIN.
FT DOMAIN 42 363 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 364 382 MEMBRANE ANCHOR.
FT DOMAIN 383 388 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 354 359 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 388 AA; 42244 MW; 3AB19EB8BCEDA722 CRC64;

Query Match 33.7%; Score 43.5; DB 1; Length 388;
Best Local Similarity 50.0%; Pred. No. 76;
Matches 12; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

OY 5 EALADQTDALQSEPAAYKADNA 28
DB 295 KALKBQL-AKQAEELAKLADKAS 317

RESULT 29
YTF3_RHILT STANDARD; PRT; 420 AA.
ID YTF3_RHILT
AC Q52871;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 45.4 kDa protein in tfuA 3' region (ORF3).
OS Rhizobium leguminosarum (biovar trifolii).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=386;
RN NCBI_TaxID=386;
RP SEQUENCE FROM N.A.
RC STRAIN=724;
RX MEDLINE=96312357; PubMed=8763943;
RA Breil B.T., Triplet E.W.;
RT "A newly discovered gene, tfuA, involved in the production of the
RT ribosomally synthesized peptide antibiotic trifollioxin.";
RL J. Bacteriol. 178:4150-4156(1996).
CC -1- SIMILARITY: BELONGS TO THE UPFR012 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement. (See <http://www.tsb-sib.ch/announce/>  
 CC or send an email to [license@tsb-sib.ch](mailto:license@tsb-sib.ch).)  
 CC -----  
 DR EMBL: U39409; AAB17514.1; -  
 DR InterPro: IPR003776; DUF181.  
 DR Pfam: PF02624; DUF181.1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 420 AA: 45372 MW: BEDDA4ED71736238 CRC64;  
 Query Match 33.7%; Score 43.5; DB 1; Length 420;  
 Best Local Similarity 34.1%; Pred. No. 83;  
 Matches 14; Conservative 7; Mismatches 7; Indels 13; Gaps 2;  
 QY 1 SDALAL-----ADQDA-----LQSEAAVVKADNNA 28  
 Db 90 SAAMEAVETRIAPADLTQATVESMRERAMIDNVA 130  
 RESULT 30  
 ID PPB\_ECOLI STANDARD; PRT; 471 AA.  
 AC P00634; P78051; P77801;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alkaline phosphatase precursor (EC 3.1.3.1) (APASE).  
 GN PHO A B0383.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JM83;  
 RX MEDLINE=8706741; PubMed=3537962;  
 RA Shuttleworth H., Taylor J., Minton N.;  
 RT "Sequence of the gene for alkaline phosphatase from *Escherichia coli*  
 JM83";  
 RL Nucleic Acids Res. 14:8689-8689(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87031576; PubMed=353724;  
 RA Chang C.N., Kuang W.-J., Chen E.Y.;  
 RT "Nucleotide sequence of the alkaline phosphatase gene of *Escherichia*  
 coli";  
 RL Gene 44:121-125(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88320572; PubMed=3045828;  
 RA Dubose R.F., Dykhuitzen D.E., Hartl D.L.;  
 RT "Genetic exchange among natural isolates of bacteria: recombination  
 within the *phoA* gene of *Escherichia coli*";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7036-7040(1988).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=9742617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 MAU B., Shao Y.;  
 RT "The complete genome sequence of *Escherichia coli* K-12";  
 RL Science 277:1453-1474(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,  
 RA Federspiel N., Hyman S., Kaiman S., Komp C., Kuri O., Lew H.,  
 RA Lin D., Nannath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBD databases.  
 RN [6]  
 RP SEQUENCE OF 1-77 FROM N.A.  
 RX MEDLINE=82081950; PubMed=6273802;  
 Kikuchi Y., Yoda K., Yamasaki M., Tamura G.;

RT "The nucleotide sequence of the promoter and the amino-terminal  
 RT region of alkaline phosphatase structural gene (*phoA*) of *Escherichia*  
 RT coli";  
 RL Nucleic Acids Res. 9:5671-5678(1981).  
 RN [7]  
 RP SEQUENCE OF 23-471 (ISOZYME 3).  
 RX MEDLINE=81273081; PubMed=7022451;  
 RA Bradshaw R.A., Cancedda F., Ericsson L.H., Neumann P.A., Piccoli S.P.,  
 RA Schlesinger M.J., Shrieffer K., Walsh K.A.;  
 RT "Amino acid sequence of *Escherichia coli* alkaline phosphatase";  
 RL Proc. Natl. Acad. Sci. U.S.A. 78:3473-3477(1981).  
 RN [8]  
 RP SEQUENCE OF 162-193 AND 385-399 FROM N.A.  
 RX MEDLINE=8211946; PubMed=7035431;  
 RA Inouye H., Barnes W., Beckwith J.;  
 RT "Signal sequence of alkaline phosphatase of *Escherichia coli*";  
 RL J. Bacteriol. 149:434-439(1982).  
 RN [9]  
 RP SEQUENCE OF 1-21 FROM N.A.  
 RX MEDLINE=86250586; PubMed=3522543;  
 RA Michaelis S., Hunt J.F., Beckwith J.;  
 RT "Effects of signal sequence mutations on the kinetics of alkaline  
 RT phosphatase export to the periplasm in *Escherichia coli*";  
 RL J. Bacteriol. 167:160-167(1986).  
 RN [10]  
 RP SEQUENCE OF 1-23 FROM N.A.  
 RX MEDLINE=89340570; PubMed=2668291;  
 RA Laforet G.A., Kaiser E.T., Kendall D.A.;  
 RT "Signal peptide subsegments are not always functionally  
 RT interchangeable. M13 procoat hydrophobic core falls to transport  
 RT alkaline phosphatase in *Escherichia coli*";  
 RL J. Biol. Chem. 264:14478-14485(1989).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE=86115281; PubMed=3910843;  
 RA Swasakul J.M., Handschumacher M.D., Krishna Murthy H.M., Foster B.A.,  
 RA Wyckoff H.W.;  
 RT "Refined structure of alkaline phosphatase from *Escherichia coli* at  
 RT 2.8-A resolution";  
 RL J. Mol. Biol. 186:417-433(1985).  
 RN [12]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=91186406; PubMed=2010919;  
 RA Kim E.E., Wyckoff H.W.;  
 RT "Reaction mechanism of alkaline phosphatase based on crystal  
 RT structures. Two-metal ion catalysis";  
 RL J. Mol. Biol. 218:449-464(1991).  
 RN [13]  
 RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS) OF MUTANT GLN-434.  
 RX MEDLINE=96194161; PubMed=8652582;  
 RA Ma L., Kantrowitz E.R.;  
 RT "Kinetic and X-ray structural studies of a mutant *Escherichia coli*  
 RT alkaline phosphatase (His-412->Gln) at one of the zinc binding  
 RT sites";  
 RL Biochemistry 35:2394-2402(1996).  
 RN [14]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE=97397347; PubMed=9253408;  
 RA Murphy J.E., Stec B., Ma L., Kantrowitz E.R.;  
 RT "Trapping and visualization of a covalent enzyme-phosphate  
 RT intermediate";  
 RL Nat. Struct. Biol. 4:618-622(1997).  
 RN [15]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
 RX MEDLINE=98202577; PubMed=9533886;  
 RA Stec B., Hehlir M.J., Brennan C., Nolte M., Kantrowitz E.R.;  
 RT "Kinetic and X-ray structural studies of three mutant *E. coli*  
 RT alkaline phosphatases: Insights into the catalytic mechanism without  
 RT the nucleophile Ser102";  
 RL J. Mol. Biol. 277:647-662(1998).  
 RN [16]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RX MEDLINE=99185045; PubMed=10085061;



RA Holtz K.M., Stec B., Kantrowitz E.R.;  
 RT "A model of the transition state in the alkaline phosphatase  
 RL J. Biol. Chem. 274:8351-8354(1999).  
 CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
 CC alcohol + phosphate.  
 CC -1- COFACTOR: METALLOENZYME CONTAINING TWO ZINC ATOMS AND A MAGNESIUM  
 CC ION.  
 CC -1- SUBUNIT: ISOZYMES 1 & 3 ARE A DIMER OF IDENTICAL CHAINS. ISOZYME 2  
 CC IS A DIMER OF HETEROGENEOUS CHAINS, ONE OF EACH OF THE SUBUNITS  
 CC FROM ISOZYMES 1 AND 3.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE ALKALINE PHOSPHATASE FAMILY.  
 CC -1- DATABASE: NAME-Prozyme technical fact sheet;  
 CC WWW="http://www.prozyme.com/technical/bap.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: V00315; CAA23603.1; ALT\_SEQ.  
 DR EMBL: X04586; CAA28257.1; -.  
 DR EMBL: AE000145; AAC73486.1; ALT\_INT.  
 DR EMBL: U73857; AAB18107.1; -.  
 DR EMBL: M29665; AAA24365.1; -.  
 DR EMBL: M29664; AAA24364.1; -.  
 DR EMBL: M33536; AAA24372.1; -.  
 DR EMBL: M13763; AAA24358.1; -.  
 DR EMBL: J05005; AAA24362.1; -.  
 DR EMBL: J01659; AAA24359.1; -.  
 DR EMBL: J01660; AAA24360.1; -.  
 DR EMBL: J01661; AAA24361.1; -.  
 DR EMBL: J01661; CAB23260.1; -.  
 DR EMBL: J01661; CAB23261.1; -.  
 DR EMBL: M13345; AAA33893.1; -.  
 DR PIR: A00776; PAECA.  
 DR PIR: A24833; A24833.  
 DR PDB: 1AJA; 14-NOV-95.  
 DR PDB: 1AJB; 14-NOV-95.  
 DR PDB: 1AJC; 14-NOV-95.  
 DR PDB: 1AJD; 14-NOV-95.  
 DR PDB: 1ALH; 27-FEB-95.  
 DR PDB: 1ALJ; 14-NOV-95.  
 DR PDB: 1ALK; 31-JAN-94.  
 DR PDB: 2ANH; 29-JAN-96.  
 DR PDB: 1ANI; 26-JAN-96.  
 DR PDB: 1AND; 26-JAN-96.  
 DR PDB: 1HQA; 08-MAR-96.  
 DR PDB: 1HJK; 15-OCT-97.  
 DR PDB: 1URR; 11-JUL-96.  
 DR PDB: 1URR; 11-JUL-96.  
 DR PDB: 1ELK; 27-MAY-98.  
 DR PDB: 1ELJ; 27-MAY-98.  
 DR PDB: 1B8J; 18-FEB-99.  
 DR ECODBASE: P046.6; 6TH EDITION.  
 DR ECGene; EG10727; phoA.  
 DR InterPro: IPR001952; Alk\_phosphatase.  
 DR Pfam: PF00245; alk\_phosphatase.  
 DR PRINTS: PR00113; ALKPHPTASE.  
 DR ProDom: PD001868; Alk\_phosphatase; 1.  
 DR SMART: SM00098; alkpc; 1.  
 DR PROSITE: PS00123; ALKALINE\_PHOSPHATASE; 1.  
 KW Hydrolase; Zinc; Magnesium; Phosphorylation; Periplasmic; Signal;  
 KW 3D-structure; Complete proteome.  
 FT CHAIN 1 21  
 FT SIGNAL 1 21  
 FT ACT\_SITE 124 124 ALKALINE PHOSPHATASE.

FT MOD\_RES 124 124 PHOSPHORYLATION.  
 Query Match 33.7%; Score 43.5; DB 1; Length 471;  
 Best Local Similarity 31.7%; Pred. No. 93;  
 Matches 13; Conservative 6; Mismatches 9; Indels 13; Caps 1;  
 QY 1 SDALFALADQIDA-----LQSEAAVYKADNNA 28  
 Db 315 NDVPTLAQMTDKAIELLSKNEKGFQVEGASIDKODHAA 355  
 RESULT 31  
 SMC3\_YEAST  
 ID SMC3\_YEAST STANDARD; PRT; 1230 AA.  
 AC P47037;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Chromosome segregation protein SMC3 (DA-box protein SMC3).  
 GN SMC3 OR YJL074C OR J1049.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=W303;  
 RX MEDLINE=97474309; PubMed=9335333;  
 RA Michaelis C., Ciosk R., Nasmyth K.;  
 RT "Cohesins: chromosomal proteins that prevent premature separation of  
 RL sister chromatids";  
 RN Cell 91:35-45(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Rose M., Koetter P., Entian K.D.;  
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Sor F.J.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE  
 CC PART OF A CHROMOSOME CONDENSATION MOTOR.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).  
 CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS  
 CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.  
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: Y14278; CAA74655.1; -.  
 DR EMBL: Z49349; CAA89366.1; -.  
 DR EMBL: X88851; CAA61313.1; -.  
 DR SGD: S0003610; SMC3.  
 DR InterPro: IPR003439; ABC\_transportr.  
 DR InterPro: IPR003405; SMC\_C.  
 DR InterPro: IPR003395; SMC\_N.  
 DR Pfam: PF02483; SMC\_C; 1.  
 DR Pfam: PF02463; SMC\_N; 1.  
 KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.  
 FT NP\_BIND 32 39  
 FT DOMAIN 172 482 COILED COIL (POTENTIAL).  
 FT DOMAIN 685 1041 COILED COIL (POTENTIAL).  
 FT SEQUENCE 1230 AA; 141336 MW; B152D88F7780341F CRC64;  
 Query Match 33.7%; Score 43.5; DB 1; Length 1230;  
 Best Local Similarity 37.0%; Pred. No. 2.5e+02;



GN UL25 OR 36.  
 OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).  
 OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicelloviruses.  
 OX NCBI\_TaxID=31520;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92295566; PubMed=1318606;  
 RA Telford E.A.R., Watson M.S., McBride K., Davidson A.J.;  
 RT "The DNA sequence of equine herpesvirus-1";  
 RL Virology 189:304-316(1992).  
 CC -1- FUNCTION: VIRION PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL25,  
 CC EHV-1 36, EBV BFRF1, HCMV UL77, ILTV ORF2, AND VZV 34.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M86664; AAB02472.1; -  
 DR PIR: B36799; M2BEC9.  
 DR InterPro: IPR002493; UL25.  
 DR Pfam: PF01499; UL25; 1.  
 SQ SEQUENCE 587 AA; 63692 MW; 6BE2F5FAB34CA3C8 CRC64;  
 QY 1 SDALBALDQTDALQSEEAAY--VKADNA 27  
 Db 98 ADALALEDAARAASADATPQVNASEA 126  
 YQYLCAEEL  
 ID YQYLCAEEL STANDARD; PRT; 621 AA.  
 AC 009293;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 69.9 kDa protein EEBD8.1 in chromosome II.  
 GN EEBD8.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Chissee S.;  
 RT Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U23484; AAC46761.1; -  
 DR WormRep: EEBD8.1; CE01882.  
 DR InterPro: IPR001097; Mat\_tudor.  
 DR Pfam: PF00567; TUDOR; 1.  
 DR SMART: SM00360; RRM; 1.  
 DR PROSITE: PS50102; RRM; 1.

DR PROSITE: PS00030; RRM\_RNP\_1; FALSE\_NEG.  
 FT Hypothetical protein: RNA-binding  
 FT DOMAIN 267 346 RNA-BINDING (RRM).  
 SQ SEQUENCE 621 AA; 69894 MW; 464DB7EF91AF65C5 CRC64;  
 QY 2 DALEBALDQTDALQSEEAAYVKADNA 27  
 Db 14 DALDIISDINGDVSSDEEAVSVKESS 39  
 YQYLCAEEL  
 ID YQYLCAEEL STANDARD; PRT; 627 AA.  
 AC 092DX9;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Chaparrone protein dnak (heat shock protein 70) (Heat shock 70 kDa  
 DE protein) (HSP70).  
 OS Dnak OR Rp185.  
 GN Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE=99039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,  
 RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria";  
 RL Nature 396:133-140(1998).  
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ235270; CA114651.1; -  
 DR HSSP: P04475; IDG4.  
 DR InterPro: IPR001023; HSP70.  
 DR Pfam: PF00012; HSP70; 1.  
 DR PRINTS: PR00301; HEATSHOCK70.  
 DR PROSITE: PS00297; HSP70\_1; 1.  
 DR PROSITE: PS00329; HSP70\_2; 1.  
 DR PROSITE: PS01036; HSP70\_3; 1.  
 KW Chaparrone: ATP-binding; Heat shock; Complete proteome.  
 SQ SEQUENCE 627 AA; 68383 MW; A638A0A5A5C8AC7 CRC64;  
 QY 5 EALADQTDALQSEEAAYVK 23  
 Db 561 EALSLKALLESDEASLIK 579  
 YQYLCAEEL  
 ID YQYLCAEEL STANDARD; PRT; 697 AA.  
 AC 092DX9;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Chaparrone protein dnak (heat shock protein 70) (Heat shock 70 kDa  
 DE protein) (HSP70).  
 OS Dnak OR Rp185.  
 GN Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE=99039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,  
 RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria";  
 RL Nature 396:133-140(1998).  
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U23484; AAC46761.1; -  
 DR WormRep: EEBD8.1; CE01882.  
 DR InterPro: IPR001097; Mat\_tudor.  
 DR Pfam: PF00567; TUDOR; 1.  
 DR SMART: SM00360; RRM; 1.  
 DR PROSITE: PS50102; RRM; 1.

```

AC P93203;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE MAR binding filament-like protein 1.
GN MPPL1.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eumsterids I; Solanales; Solanaceae; Solanum.
CX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. VENT CHERRY; TISSUE=fruit;
RX MEDLINE=97112038; PubMed=8953774;
RA Meier I., Phelan T., Grunsem W., Spiker S., Schneider D.;
RT "MPPL, a novel plant filament-like protein with affinity for matrix
RT attachment region DNA."
RL Plant Cell 8:2105-2115(1996).
CC -1- FUNCTION: BINDS DNA. INTERACTS WITH CHROMATIN VIA MATRIX
CC ATTACHMENT REGIONS (MARS). LIKELY TO PARTICIPATE IN NUCLEAR
CC ARCHITECTURE BY CONNECTING CHROMATIN WITH THE NUCLEAR MATRIX AND
CC POTENTIALLY WITH THE NUCLEAR ENVELOPE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR MATRIX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Y07861; CAA69181.1;
KM Nuclear protein: DNA-binding; Coiled coil.
FT DOMAIN 125 681 COILED COIL (POTENTIAL).
FT DOMAIN 10 15 POLY-SER.
SQ SEQUENCE 697 AA; 79516 MW; 700A36D68D6A7E49 CRC64;

Query Match 33.3%; Score 43; DB 1; Length 697;
Best Local Similarity 43.5%; Pred. NO. 1.6e+02;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 SCALEALADQTDALQSEBAVVK 23
DB 508 SDELVAVSENDSLOTETLVYVK 530

RESULT 38
ID METE_MYCTU STANDARD: PRT: 759 AA.
AC 006584;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE 5-methyltetrahydropteroyltri-glutamate--homocysteine methyltransferase
DE (EC 2.1.1.14) (Methionine synthase, vitamin-B12 independent isozyme)
DE (Cobalamin-independent methionine synthase).
GN METE OR RV1133C OR MT1165 OR MTC2268.22.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=968295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

```

```

RA Ruter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischman R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mkura A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE TRANSFER OF A METHYL GROUP FROM 5-
CC METHYLTERAHYDROFOLATE TO HOMOCYSTEINE RESULTING IN METHIONINE
CC FORMATION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 5-methyltetrahydropteroyl-L-glutamate + L-
CC homocysteine = tetrahydropteroyl-L-glutamate + L-methionine.
CC -1- COFACTOR: ZINC. BINDS ONE ION PER SUBUNIT (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.
CC -1- SIMILARITY: BELONGS TO THE VITAMIN-B12 INDEPENDENT METHIONINE
CC SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z95585; CAB09044.1;
DR EMBL; AE006955; AAK45422.1;
DR TIGR; MT1165;
DR TuberculList; RV1133C;
DR InterPro: IPR002629; Methionine_synth.
DR Pfam: PF01717; Methionine_synth; 1.
KM Transferase: Methyltransferase; Methionine biosynthesis; Zinc; Repeat;
KW Complete proteome.
FT METAL 647 647 ZINC (BY SIMILARITY).
FT METAL 649 649 ZINC (BY SIMILARITY).
FT METAL 732 732 ZINC (BY SIMILARITY).
SQ SEQUENCE 759 AA; 81581 MW; 702F90BF79B25C8D CRC64;

Query Match 33.3%; Score 43; DB 1; Length 759;
Best Local Similarity 43.5%; Pred. NO. 1.8e+02;
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 6 ALADQTDALQSEBAVVKADNAA 28
DB 586 AIRDETVDLQSGAVIQVDEPA 608

RESULT 39
ID YB52_MYCPN STANDARD: PRT: 794 AA.
AC P75034;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein MPN152 precursor (E07_orf794).
GN MPN152 OR MP002.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;

```

RA Himmelreich R., Hilbert H., Plagens H., Plöckl E., Li B.-C.,  
 RA Hermann R.;  
 RT \*Complete sequence analysis of the genome of the bacterium Mycoplasma  
 RT pneumoniae.\*;  
 CC Nucleic Acids Res. 24:4420-4449(1996).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor  
 CC (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE MG185 / MG260 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: A600001; AAB95650.1; -;  
 DR Hypothetical protein; Lipoprotein; Membrane; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 22  
 FT CHAIN 23 794  
 FT LIPID 23 23  
 FT SEQUENCE 794 AA; 86871 MW; 0DA72D12807F53EE CRC64;  
 SQ  
 Query Match 33.3%; Score 43; DB 1; Length 794;  
 Best Local Similarity 45.5%; Pred. No. 1.9e+02; Mismatches 8; Indels 0; Gaps 0;  
 Matches 10; Conservative 4;  
 QY 3 ALBALDQTDALQSEEAAYVKA 24  
 Db 634 ATETETDKSNTLSQNEAIVLPA 655  
 ID MML5\_MYCTU STANDARD; PRT; 964 AA.  
 AC 053784;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative membrane protein mmpL5.  
 GN MML5 OR RV0676C OR MT0705 OR MTV040.04C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.\*;  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwin M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT \*Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.\*;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE MML FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL021943; CAAL7459.1; -;  
 DR EMBL: A6006964; AAK44930.1; -;  
 DR TIGR: MT0705; -;  
 DR Tuberculosis; RV0676c; -;  
 DR InterPro: IPR001036; ACR\_tran.  
 DR PRINTS: PR00702; ACRIFLAVINRP.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 31 51  
 FT TRANSMEM 203 223  
 FT TRANSMEM 230 250  
 FT TRANSMEM 255 275  
 FT TRANSMEM 302 322  
 FT TRANSMEM 340 360  
 FT TRANSMEM 389 409  
 FT TRANSMEM 745 765  
 FT TRANSMEM 774 794  
 FT TRANSMEM 803 823  
 FT TRANSMEM 826 846  
 FT TRANSMEM 880 900  
 FT TRANSMEM 901 921  
 FT CONFLICT 948 948  
 SQ SEQUENCE 964 AA; 104784 MW; B7C945940A1176BD CRC64;  
 I -> V (IN REF. 2).  
 Query Match 33.3%; Score 43; DB 1; Length 964;  
 Best Local Similarity 41.7%; Pred. No. 2.3e+02;  
 Matches 10; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
 QY 5 BALADQTDALQSEEAAYVKAADNA 28  
 Db 660 EGIQDQMAAQMDESAANGAEAFDAS 683

Search completed: October 13, 2002, 04:46:24  
 Job time : 5.04278 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus.p2n model

Run on: October 13, 2002, 04:54:53 ; Search time 8.98396 Seconds  
(without alignments)  
765.558 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALFALADQTDALQSEEAAYVADNAA 28

Scoring table: BLOSUM62  
Xgapop 10.0 , Ygapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODEL=frame.p2n.model -DEV=x1h  
-O=/cg2\_1/USPTO.spool/US09847539/runtat\_10102002\_093106\_5020/app.query.fasta-1.526  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=std.rni -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US09847539.ecgn\_1.1.13-enumat\_10102002\_093106\_5020 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-MAIN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Issued\_Patents\_NA.\*  
1: /cg2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cg2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cg2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cg2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cg2\_6/ptodata/2/ina/PCBUS.COMB.seq.\*  
6: /cg2\_6/ptodata/2/ina/backfile1.seq.\*

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length DB	ID	Description
C 1	47	36.4	59065	4	US-09-813-817-3
C 2	47	36.4	4403765	4	US-09-103-840A-2
C 3	47	36.4	4411529	4	US-09-103-840A-1
C 4	45	34.9	4403765	4	US-09-103-840A-2
C 5	45	34.9	4411529	4	US-09-103-840A-1
C 6	44.5	34.5	1548	2	US-08-762-106-5
C 7	44.5	34.5	1548	2	US-09-320-774-5
C 8	44.5	34.5	1581	2	US-08-762-106-6
C 9	44.5	34.5	1581	2	US-09-320-774-6
C 10	44	34.1	1033	3	US-08-961-083-191
C 11	44	34.1	1772	2	US-08-960-022-13
C 12	43.5	33.7	744	2	US-08-743-637B-19

C 13	43.5	33.7	744	3	US-08-526-840B-19	Sequence 19, Appl
C 14	43.5	33.7	1149	1	US-08-049-282B-1	Sequence 1, Appl1
C 15	43.5	33.7	1149	1	US-08-537-715-1	Sequence 1, Appl1
C 16	43.5	33.7	1149	5	PCT-US94-04173-1	Sequence 1, Appl1
C 17	43.5	33.7	1454	2	US-08-657-392-1	Sequence 1, Appl1
C 18	43.5	33.7	1454	5	PCT-US94-02539-1	Sequence 1, Appl1
C 19	43.5	33.7	1455	2	US-08-657-392-26	Sequence 26, Appl
C 20	43.5	33.7	1455	5	PCT-US94-02539-26	Sequence 26, Appl
C 21	43.5	33.7	1857	5	US-08-537-715-3	Sequence 3, Appl1
C 22	43.5	33.7	1857	5	PCT-US94-04173-3	Sequence 3, Appl1
C 23	43.5	33.7	2233	1	US-08-235-838-6	Sequence 6, Appl1
C 24	43.5	33.7	2233	4	US-08-465-473B-6	Sequence 6, Appl1
C 25	43.5	33.7	2681	2	US-08-928-213B-7	Sequence 7, Appl1
C 26	43.5	33.7	6162	2	US-08-673-312-9	Sequence 9, Appl1
C 27	43	33.3	1221	2	US-08-934-846-3	Sequence 3, Appl1
C 28	43	33.3	1221	4	US-09-238-557-3	Sequence 3, Appl1
C 29	43	33.3	1254	2	US-08-934-846-1	Sequence 1, Appl1
C 30	43	33.3	1254	4	US-09-238-557-1	Sequence 1, Appl1
C 31	43	33.3	4224	1	US-08-612-521-1	Sequence 1, Appl1
C 32	42.5	32.9	1274	2	US-08-470-670A-1	Sequence 1, Appl1
C 33	42.5	32.9	1274	4	US-08-461-511A-1	Sequence 1, Appl1
C 34	42.5	32.9	1274	5	PCT-US94-07089-1	Sequence 1, Appl1
C 35	42.5	32.9	1303	1	US-08-081-448-1	Sequence 1, Appl1
C 36	42.5	32.9	43280	2	US-08-804-227C-1	Sequence 1, Appl1
C 37	42	32.6	621	2	US-08-943-915-4	Sequence 4, Appl1
C 38	42	32.6	927	4	US-09-254-733-4	Sequence 1, Appl1
C 39	42	32.6	1060	1	US-08-090-013-1	Sequence 1, Appl1
C 40	42	32.6	1060	1	US-08-081-328-1	Sequence 1, Appl1
C 41	42	32.6	1060	1	US-08-232-249-1	Sequence 1, Appl1
C 42	42	32.6	1060	2	US-08-921-426-7	Sequence 7, Appl1
C 43	42	32.6	1060	2	US-08-833-642A-1	Sequence 1, Appl1
C 44	42	32.6	1060	2	US-08-140-008A-3	Sequence 3, Appl1
C 45	42	32.6	1060	2	US-08-389-423-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-09-813-817-3/c  
: Sequence 3, Application US/09813817  
: Patent No. 6340583  
: GENERAL INFORMATION:

: APPLICANT: YAN, Chunhua et al.  
: TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
: FILE REFERENCE: CLO01178  
: CURRENT FILING DATE: 2001-03-22  
: NUMBER OF SEQ ID NOS: 4  
: SOFTWARE: FastSeq for Windows Version 4.0  
: SEQ ID NO 3  
: LENGTH: 59065  
: TYPE: DNA  
: ORGANISM: Human  
US-09-813-817-3

Alignment Scores:  
Pred. No.: 2.7e+03  
Score: 47.00  
Percent Similarity: 76.47%  
Best Local Similarity: 58.82%  
Query Match: 36.43%  
DB: 4  
Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-09-813-817-3 (1-59065)  
QY 8 AlaApgIntRpaAlaLeuGlnSerGluGuaAlaValValysAla 24  
DB 21239 GCTGAACCAACTGCTCTAGAACACGCTGACGCTTCATAAGCG 21189

RESULT 2  
US-09-103-840A-2/c





ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37RV  
US-09-103-840A-1

## Alignment Scores:

Pred. No.:	6.3e+05	Length:	4411529
Score:	45.00	Matches:	13
Percent Similarity:	55.88%	Conservative:	6
Best Local Similarity:	38.24%	Mismatches:	7
Query Match:	34.88%	Indels:	8
DB:	4	Gaps:	1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-103-840A-1 (1-4411529)

OY 3 ALaLeuGuaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVal 22

DB 1667153 GCCAACGACGCCCTGCAAGACCTGACGAGCTTCAGCGCGAAGAAAGCGTTAAAC 1667212

OY 23 LysAla-----AspAsnAlaAla 28

DB 1667213 AAGCGCATGTCGACGTGGAAACCGCTCGGACAAACGCTGCC 1667254

## RESULT 6

US-08-762-106-5  
Sequence 5, Application US/08762106  
Patent No. 5948677

## GENERAL INFORMATION:

APPLICANT: Jarvik, Jonathan W.  
TITLE OF INVENTION: READING FRAME INDEPENDENT EPIOTOPE  
SOFTWARE: PatentID Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: 47  
FILING DATE: 09-DEC-1996

## CORRESPONDENCE ADDRESS:

ADDRESSER: Harris Brotman  
STREET: 202 Coast Blvd., Suite 111  
CITY: La Jolla  
STATE: California  
COUNTRY: US  
ZIP: 92037

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentID Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/762,106  
FILING DATE: 09-DEC-1996  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Brotman, Harris F.  
REGISTRATION NUMBER: 35,461  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 654-2428  
INFORMATION FOR SEQ ID NO: 5:

## SEQUENCE CHARACTERISTICS:

LENGTH: 1548 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-762-106-5

## Alignment Scores:

Pred. No.:	76.7	Length:	1548
Score:	44.50	Matches:	15
Percent Similarity:	65.52%	Conservative:	4
Best Local Similarity:	51.72%	Mismatches:	7
Query Match:	34.50%	Indels:	3
DB:	2	Gaps:	2

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-762-106-5 (1-1548)

OY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 20

DB 1105 GCCGATGCCGTGAGGCGGCGCGCGGAGCTGACGCGC---CAGCGGAGGAGGCGCGCC 1161

OY 21 ValValLys-----AlaAspAsnAla 27

DB 1162 AACGCCAAGTGGAGGCGGACAAAGCGC 1188

## RESULT 7

US-09-320-774-5  
Sequence 5, Application US/09320774  
Patent No. 6265545

## GENERAL INFORMATION:

APPLICANT: Jarvik, Jonathan W.  
TITLE OF INVENTION: READING FRAME INDEPENDENT EPIOTOPE  
SOFTWARE: PatentID Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: 47  
FILING DATE: 09-DEC-1996

## CORRESPONDENCE ADDRESS:

ADDRESSER: Harris Brotman  
STREET: 202 Coast Blvd., Suite 111  
CITY: La Jolla  
STATE: California  
COUNTRY: US  
ZIP: 92037

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentID Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/320,774  
FILING DATE: 09-DEC-1996

## CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/762,106  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Brotman, Harris F.  
REGISTRATION NUMBER: 35,461  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 654-2428  
INFORMATION FOR SEQ ID NO: 5:

## SEQUENCE CHARACTERISTICS:

LENGTH: 1548 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-320-774-5

## Alignment Scores:

Pred. No.:	76.7	Length:	1548
Score:	44.50	Matches:	15
Percent Similarity:	65.52%	Conservative:	4
Best Local Similarity:	51.72%	Mismatches:	7
Query Match:	34.50%	Indels:	3
DB:	4	Gaps:	2

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-320-774-5 (1-1548)

OY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 20

DB 1105 GCCGATGCCGTGAGGCGGCGCGCGGAGCTGACGCGC---CAGCGGAGGAGGCGCGCC 1161

OY 21 ValValLys-----AlaAspAsnAla 27

DB 1162 AACGCCAAGTGGAGGCGGACAAAGCGC 1188

## RESULT 8

US-08-762-106-6  
Sequence 6, Application US/08762106

```
? Patent No. 5948677
? GENERAL INFORMATION:
? APPLICANT: Jarvik, Jonathan W.
? TITLE OF INVENTION: READING FRAME INDEPENDENT EPTIPOE
? TITLE OF INVENTION: TAGGING
? NUMBER OF SEQUENCES: 47
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Harris Brotman
? STREET: 202 Coast Blvd., Suite 111
? CITY: La Jolla
? STATE: California
? COUNTRY: US
? ZIP: 92037
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/762,106
? FILING DATE: 09-DEC-1996
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Brotman, Harris F.
? REGISTRATION NUMBER: 35,461
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (619) 654-2428
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1581 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
? US-08-762-106-6

Alignment Scores:
Pred. No.: 78.7 Length: 1581
Score: 44.50 Matches: 15
Percent Similarity: 65.52% Conservative: 4
Best Local Similarity: 51.72% Mismatches: 7
Query Match: 34.50% Indels: 3
DB: 2 Gaps: 2

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-762-106-6 (1-1581)
QY 1 SetAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
Db 1138 GCCGATCCGCTCGAGCGCGCGGCGGCGGAGCTGACGCGC---CAGCGCGAGAGCGCGGC 1194
QY 21 ValValIys-----AlaAspAsnAla 27
Db 1195 AACGCCAAGTGGAGCGCGGACAAAGGCG 1221

RESULT 9
US-09-320-774-6
? Sequence 6, Application US/09320774
? Patent No. 6255545
? GENERAL INFORMATION:
? APPLICANT: Jarvik, Jonathan W.
? TITLE OF INVENTION: READING FRAME INDEPENDENT EPTIPOE
? TITLE OF INVENTION: TAGGING
? NUMBER OF SEQUENCES: 47
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Harris Brotman
? STREET: 202 Coast Blvd., Suite 111
? CITY: La Jolla
? STATE: California
? COUNTRY: US
? ZIP: 92037
? COMPUTER READABLE FORM:
```

```
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/320,774
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/762,106
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Brotman, Harris F.
? REGISTRATION NUMBER: 35,461
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (619) 654-2428
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1581 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
? US-09-320-774-6

Alignment Scores:
Pred. No.: 78.7 Length: 1581
Score: 44.50 Matches: 15
Percent Similarity: 65.52% Conservative: 4
Best Local Similarity: 51.72% Mismatches: 7
Query Match: 34.50% Indels: 3
DB: 2 Gaps: 2

US-09-847-539a-6_COPY_59_86 (1-28) x US-09-320-774-6 (1-1581)
QY 1 SetAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
Db 1138 GCCGATCCGCTCGAGCGCGCGGCGGCGGAGCTGACGCGC---CAGCGCGAGAGCGCGGC 1194
QY 21 ValValIys-----AlaAspAsnAla 27
Db 1195 AACGCCAAGTGGAGCGCGGACAAAGGCG 1221

RESULT 10
US-09-961-083-191
? Sequence 191, Application US/08961083
? Patent No. 6159469
? GENERAL INFORMATION:
? APPLICANT: Chol et. al.
? TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
? NUMBER OF SEQUENCES: 452
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Human Genome Sciences, Inc.
? STREET: 9410 Key West Avenue
? CITY: Rockville
? STATE: Maryland
? COUNTRY: USA
? ZIP: 20850
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
? COMPUTER: HP Vectra 486/33
? OPERATING SYSTEM: MSDOS version 6.2
? SOFTWARE: ASCII Text
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/961,083
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
```

NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 191:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1033 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-083-191

Alignment Scores:  
Pred. No.: 56.4 Length: 1033  
Score: 44.00 Matches: 11  
Percent Similarity: 65.00% Conservative: 2  
Best local Similarity: 55.00% Mismatches: 7  
Query Match: 34.11% Indels: 0  
DB: 3 gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-961-083-191 (1-1033)

OY 8 AlaspclnThraspalaleuInserGlunlaAlaValLysAlaAspaNaLa 27  
DB 8 GCAGGCGACAGATGCTCGCAATTGAAAGGCGAGTACCGAAGGAAAAAGCA 67

RESULT 11  
US-08-960-022-13  
Sequence 13, Application US/08960022  
Patent No. 5976837  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallee, Edward R.  
APPLICANT: Racine, Lisa A.  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/960,022  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1772 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

US-08-960-022-13

Alignment Scores:  
Pred. No.: 111 Length: 1772  
Score: 44.00 Matches: 9  
Percent Similarity: 68.00% Conservative: 8  
Best local Similarity: 36.00% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 2 gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-960-022-13 (1-1772)

OY 3 AlaleuGlAlaleuAlaAspGlThraspalaleuInserGlunlaAlaVal 22  
DB 795 AGCATCAAGCGCAAGAGCGACAGCTGCTGAAGAGAGAGAGCGCGCTGTG 854

OY 23 LysAlaAspaNaLa 27  
DB 855 GAAAGTAACAACGCA 869

RESULT 12  
US-08-743-637B-19/C  
Sequence 19, Application US/08743637B  
Patent No. 5994066  
GENERAL INFORMATION:  
APPLICANT: BERGERON, Michel G.  
APPLICANT: PICARD, Francois J.  
APPLICANT: OUELLETTE, Marc  
APPLICANT: ROY, Paul H.  
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA  
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND  
IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED  
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OUARLES & BRADY  
STREET: 411 EAST WISCONSIN AVENUE  
CITY: MILWAUKEE  
STATE: WISCONSIN  
COUNTRY: USA  
ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/743,637B  
FILING DATE: 04-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/526,840  
FILING DATE: 11-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BAKER, Jean C.  
REGISTRATION NUMBER: 35,433  
REFERENCE/DOCKET NUMBER: 850586,90012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5000  
TELEFAX: (414) 277-5591  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 744 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: Pseudomonas aeruginosa  
US-08-743-637B-19

Alignment Scores:  
Pred. No.: 45.7 Length: 744





```

; APPLICANT: Yue, V.T.
; TITLE OF INVENTION: Genetically Engineered Enzymes And Their
; TITLE OF INVENTION: Conjugates For Diagnostic Assays
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: SoftPC
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/657,392
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/100,708
; FILING DATE: July 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Wong, Wean Khing
; REGISTRATION NUMBER: 33,561
; REFERENCE/DOCKET NUMBER: 5324.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (708) 938-3517
; TELEFAX: (708) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1454 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
;
US-08-657-392-1
;
Alignment Scores:
Pred. No.: 106 Length: 1454
Score: 43.50 Matches: 13
Percent Similarity: 46.34% Conservative: 6
Best Local Similarity: 31.71% Mismatches: 9
Query Match: 33.72% Indels: 13
DB: Gaps: 1

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-657-392-1 (1-1454)
OY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAla----- 13
DB 968 AACGACCTGGTTCCAAACCCCTGGCGCAGATGACCGCAAAAGCCATTGAACCTTTAAGTAAA 1027
OY 14 -----LeuGlnSerGluGluAlaAlaValAlaValAlaAspAsnAla 27
DB 1028 AATGAGAAAGCGTTTCTCTCGCAAGTTGAAGGTGGCTCATGATGATTAACAGATCATGCT 1087
OY 28 Ala 28
DB 1088 GCG 1090

RESULT 18
PCT-US94-02539-1
; Sequence 1, Application PC/TUS9402539
; GENERAL INFORMATION:
; APPLICANT: Brate, E.M.
; APPLICANT: Brennan, C.A.
; APPLICANT: Bridon, D.P.
; APPLICANT: Jaffe, K.D.
; APPLICANT: Kraft, G.A.
```

```

; APPLICANT: Mandeckl, W.
; APPLICANT: March, S.C.
; APPLICANT: Russell, J.R.
; APPLICANT: Yue, V.T.
; TITLE OF INVENTION: Genetically Engineered Enzymes
; TITLE OF INVENTION: And Their
; TITLE OF INVENTION: Conjugates For Diagnostic Assays
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: SoftPC
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/02539
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Wong, Wean Khing
; REGISTRATION NUMBER: 33,561
; REFERENCE/DOCKET NUMBER: 5324.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (708) 938-3517
; TELEFAX: (708) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1454 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
;
PCT-US94-02539-1
;
Alignment Scores:
Pred. No.: 106 Length: 1454
Score: 43.50 Matches: 13
Percent Similarity: 46.34% Conservative: 6
Best Local Similarity: 31.71% Mismatches: 9
Query Match: 33.72% Indels: 13
DB: Gaps: 1

US-09-847-539a-6_COPY_59_86 (1-28) x PCT-US94-02539-1 (1-1454)
OY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAla----- 13
DB 968 AACGACCTGGTTCCAAACCCCTGGCGCAGATGACCGCAAAAGCCATTGAACCTTTAAGTAAA 1027
OY 14 -----LeuGlnSerGluGluAlaAlaValAlaValAlaAspAsnAla 27
DB 1028 AATGAGAAAGCGTTTCTCTCGCAAGTTGAAGGTGGCTCATGATGATTAACAGATCATGCT 1087
OY 28 Ala 28
DB 1088 GCG 1090

RESULT 19
US-08-657-392-26
; Sequence 26, Application US/08657392
; Patent No. 5843634
; GENERAL INFORMATION:
; APPLICANT: Brate, E.M.
; APPLICANT: Brennan, C.A.
; APPLICANT: Bridon, D.P.
```

APPLICANT: Jaffe, K.D.  
APPLICANT: Kraft, G.A.  
APPLICANT: Mandeckl, W.  
APPLICANT: March, S.C.  
APPLICANT: Russell, J.R.  
APPLICANT: Yue, V.T.  
TITLE OF INVENTION: Genetically Engineered Enzymes And Their  
TITLE OF INVENTION: Conjugates For Diagnostic Assays  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES  
STREET: One Abbott Park Road  
CITY: Abbott Park  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: SoftPC  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/657,392  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/100,708  
FILING DATE: July 29, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Wong, Wean Khing  
REGISTRATION NUMBER: 33,561  
REFERENCE/DOCKET NUMBER: 5324.US.P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 938-3517  
TELEFAX: (708) 938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1455 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE:  
ORIGINAL SOURCE:  
ORGANISM:  
US-08-657-392-26

Alignment Scores:  
Pred. No.: 106 Length: 1455  
Score: 43.50 Matches: 13  
Percent Similarity: 46.34% Conservative: 6  
Best Local Similarity: 31.71% Mismatches: 9  
Query Match: 33.72% Indels: 13  
Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-657-392-26 (1-1455)

QY 1 SeraspAlaLeuGluAlaLeuAlaAspGlnThrAspAla----- 13  
Db 969 AACGACCTGGTTCCAAACCTGGCGAGATGACCAAGCCATTGAACCTTTAAGTAA 1028  
QY 14 -----LeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 27  
Db 1029 AATGAAAGAGCTTTTCTCTCAAGTTGAAGTGGCTCAATCGATTAACAGATCATCT 1088  
QY 28 Ala 28  
Db 1089 GCG 1091

RESULT 20  
PCT-US94-02539-26  
Sequence 26, Application PC/TUS9402539  
GENERAL INFORMATION:

APPLICANT: Brate, E.M.  
APPLICANT: Brennan, C.A.  
APPLICANT: Bridon, D.P.  
APPLICANT: Jaffe, K.D.  
APPLICANT: Kraft, G.A.  
APPLICANT: Mandeckl, W.  
APPLICANT: March, S.C.  
APPLICANT: Russell, J.R.  
APPLICANT: Yue, V.T.  
TITLE OF INVENTION: Genetically Engineered Enzymes  
TITLE OF INVENTION: And Their  
TITLE OF INVENTION: Conjugates For Diagnostic Assays  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES  
STREET: One Abbott Park Road  
CITY: Abbott Park  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: SoftPC  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/02539  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wong, Wean Khing  
REGISTRATION NUMBER: 33,561  
REFERENCE/DOCKET NUMBER: 5324.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 938-3517  
TELEFAX: (708) 938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1455 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE:  
ORIGINAL SOURCE:  
ORGANISM:  
PCT-US94-02539-26

Alignment Scores:  
Pred. No.: 106 Length: 1455  
Score: 43.50 Matches: 13  
Percent Similarity: 46.34% Conservative: 6  
Best Local Similarity: 31.71% Mismatches: 9  
Query Match: 33.72% Indels: 13  
Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x PCT-US94-02539-26 (1-1455)

QY 1 SeraspAlaLeuGluAlaLeuAlaAspGlnThrAspAla----- 13  
Db 969 AACGACCTGGTTCCAAACCTGGCGAGATGACCAAGCCATTGAACCTTTAAGTAA 1028  
QY 14 -----LeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 27  
Db 1029 AATGAAAGAGCTTTTCTCTCAAGTTGAAGTGGCTCAATCGATTAACAGATCATCT 1088  
QY 28 Ala 28  
Db 1089 GCG 1091

RESULT 21  
US-08-537-715-3  
Sequence 3, Application US/08537715

```

: Patent No. 5910627
: GENERAL INFORMATION:
: APPLICANT: Chuck, George S.
: APPLICANT: Dooner, Hugo K.
: APPLICANT: Courtney-Guterson, Neal
: APPLICANT: Keller, Janis
: APPLICANT: Nijjar, Charnjit S.
: APPLICANT: Ralston, Edward J.
: TITLE OF INVENTION: PH Genes and Their Uses
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/537,715
: FILING DATE: 16-FEB-1996
: CLASSIFICATION: 800
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/049,282
: FILING DATE: 16-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US94/04173
: FILING DATE: 15-APR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Bastian, Kevin L.
: REGISTRATION NUMBER: 34,774
: REFERENCE/DOCKET NUMBER: 012176-003410US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1857 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 481..1632
: OTHER INFORMATION: /note="Ph6 gene CDNA"
: US-08-537-715-3

Alignment Scores:
Pred. No.: 144 Length: 1857
Score: 43.50 Matches: 11
Percent Similarity: 76.19% Conservative: 5
Best Local Similarity: 52.38% Mismatches: 4
Query Match: 33.72% Indels: 1
DB: 2 Gaps: 1

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-537-715-3 (1-1857)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAla---LeuGlnSerGluGluAlaAlaValVal 22
DB 1210 CTTGAAGCTTAGACCAATCAGACGAGGCTACCGCTCCAGACAAAGATACAGGCTACTGTG 1269
QY 23 Lys 23
DB 1270 AAG 1272

RESULT 22
PCT-US94-04173-3
: Sequence 3, Application PC/TUS9404173
```

```

: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: PH GENES AND THEIR USES
: NUMBER OF SEQUENCES: 4
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04173
: FILING DATE: 15-APR-1994
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/049,282
: FILING DATE: 16-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Bastian, Kevin L.
: REGISTRATION NUMBER: 34,774
: REFERENCE/DOCKET NUMBER: 12176-34-1PC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-5043
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1857 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 481..1632
: PCT-US94-04173-3

Alignment Scores:
Pred. No.: 144 Length: 1857
Score: 43.50 Matches: 11
Percent Similarity: 76.19% Conservative: 5
Best Local Similarity: 52.38% Mismatches: 4
Query Match: 33.72% Indels: 1
DB: 5 Gaps: 1

US-09-847-539a-6_COPY_59_86 (1-28) x PCT-US94-04173-3 (1-1857)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAla---LeuGlnSerGluGluAlaAlaValVal 22
DB 1210 CTTGAAGCTTAGACCAATCAGACGAGGCTACCGCTCCAGACAAAGATACAGGCTACTGTG 1269
QY 23 Lys 23
DB 1270 AAG 1272

RESULT 23
US-08-235-838-6
: Sequence 6, Application US/08235838
: Patent No. 5571894
: GENERAL INFORMATION:
: APPLICANT: Wels, Winfried S.
: APPLICANT: Hynes, Nancy E.
: APPLICANT: Harwerth, Ina-Maria
: APPLICANT: Groner, Bernd
: APPLICANT: Hardman, No. 5571894man
: APPLICANT: Zwickl, Markus
: TITLE OF INVENTION: Recombinant Antibodies Specific for a
: TITLE OF INVENTION: Growth Factor Receptor
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: New York
: COUNTRY: USA
```



ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/235, 838  
FILING DATE: TBA  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/828, 832  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 91-810079.3  
FILING DATE: 05-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2233 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Mouse and E. coli  
INDIVIDUAL ISOLATE: E. coli  
IMMEDIATE SOURCE:  
CLONE: pMW616  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 23..2155 /note= "89-445 FRP5 heavy chain  
OTHER INFORMATION: var domain; 446-490 is a linker sequence; 491-814  
OTHER INFORMATION: FRP5 light chain var domain; 815-2155 coding region  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..22  
OTHER INFORMATION: /function= "ompa 5'non-coding  
OTHER INFORMATION: region"  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 23..85  
OTHER INFORMATION: /note= "ompa signal peptide"  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 2156..2233  
OTHER INFORMATION: /function= "phoa 3' non-coding  
OTHER INFORMATION: region"  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 86..2155  
US-08-235-838-6  
Alignment Scores:  
Pred. No.: 181 Length: 2233  
Score: 43.50 Matches: 13  
Percent Similarity: 46.34% Conservative: 6  
Best Local Similarity: 31.71% Mismatches: 9  
Query Match: 33.72% Indels: 13  
DB: 1 Gaps: 1  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-235-838-6 (1-2233)

QY 1 SeraspalaLeuglAlaLeuAlaAspGlnThraspala----- 13  
Db 1685 AATGACAGTGTACCAACCGCTGGCCAGATGACGACAAAGCCATTGATGTGAGTAA 1744  
QY 14 -----LeuGlnSerGluGluAlaValValLysAlaAspAsnAla 27  
Db 1745 AATGAGAAGGCTTTTCCCTGCAAGTTGAAGTGCATCAATGATTAACAGATCATGCT 1804  
QY 28 Ala 28  
Db 1805 GCG 1807  
RESULT 24  
US-08-465-4738-6  
Sequence 6, Application US/08465473B  
Patent No. 5939531  
GENERAL INFORMATION:  
APPLICANT: Wels, Manfred S.  
APPLICANT: Hynes, Nancy E.  
APPLICANT: Harwerth, Ina-Maria  
APPLICANT: Groner, Bernd  
APPLICANT: Hardman, No. 5939531man  
APPLICANT: Zwickl, Markus  
TITLE OF INVENTION: Recombinant Antibodies Specific for a  
TITLE OF INVENTION: Growth Factor Receptor  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NOVARTIS Corporation  
STREET: 564 Morris Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901-6940  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,473B  
FILING DATE: 5 June 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/828, 832  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 91-810079.3  
FILING DATE: 05-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Pfeiffer, Henna J.  
REGISTRATION NUMBER: 22,640  
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)522 6940  
TELEFAX: (908)522 6955  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2233 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Mouse and E. coli  
INDIVIDUAL ISOLATE: E. coli  
IMMEDIATE SOURCE:  
CLONE: pMW616  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 23..2155  
OTHER INFORMATION: /note= "89-445 FRP5 heavy chain

OTHER INFORMATION: var.domain: 446-490 15 aa linker sequence; 491-814  
OTHER INFORMATION: FRP5 light chain var.domain: 815-2155 coding region  
OTHER INFORMATION: of phoa  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..22  
OTHER INFORMATION: /function="ompa 5' non-coding region"  
FEATURE:  
NAME/KEY: s19-peptide  
LOCATION: 23..85  
OTHER INFORMATION: /note="ompa signal peptide"  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 2156..2233  
OTHER INFORMATION: /function="phoa 3' non-coding region"  
OTHER INFORMATION: region"  
FEATURE:  
NAME/KEY: mat-peptide  
LOCATION: 86..2155  
US-08-465-4738-6

Alignment Scores:  
Pred. No.: 181 Length: 2233  
Score: 43.50 Matches: 13  
Percent Similarity: 46.348 Conservative: 9  
Best Local Similarity: 31.718 Mismatches: 6  
Query Match: 33.728 Indels: 13  
DB: 2 Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-465-4738-6 (1-2233)  
QY 1 SerAspAlaLeuGluaLalaAspGlnThrAspAla----- 13  
Db 1685 AATGACAGTGTACCAACCTGGCGGCAGATGACGACAAACCAATTGATTTGTCAGTAA 1744  
QY 14 -----LeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 27  
Db 1745 AATGAGAAAGCTTTTCCTGCAAGCTGCTCATGATTAACAGATCATGCT 1804  
QY 28 Ala 28  
Db 1805 GCG 1807

RESULT 25  
US-08-928-2138-7  
Sequence 7, Application US/089282138  
Patent No. 6238905  
GENERAL INFORMATION:  
APPLICANT: McHenry, Charles S.  
Seville, Mark  
Gull, Millard G.  
TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III  
NUMBER OF SEQUENCES: 195  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,2138  
FILING DATE: 12-Sep-1997  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: MacKnight, Kamrin T.

REGISTRATION NUMBER: 38,230  
REFERENCE/DOCKET NUMBER: ENZCO-02550  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-705-8410  
TELEFAX: 415-397-8338  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2681 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-08-928-2138-7

Alignment Scores:  
Pred. No.: 228 Length: 2681  
Score: 43.50 Matches: 12  
Percent Similarity: 73.08 Conservative: 7  
Best Local Similarity: 46.158 Mismatches: 6  
Query Match: 33.728 Indels: 1  
DB: 4 Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-928-2138-7 (1-2681)  
QY 2 AspAlaLeuGluaLalaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21  
Db 1686 GAGGCCATGAGAGGCGCTCGCCGCCGCGCTCCGACGCCCTTA---AGCCTGAGAGTGCGCCCTC 1742  
QY 22 ValLysAlaAspAsnAla 27  
Db 1743 CTGAGCGCGGAGAGGCC 1760

RESULT 26  
US-08-673-312-9  
Sequence 9, Application US/08673312  
Patent No. 5891699  
GENERAL INFORMATION:  
APPLICANT: BOULATIN, JEAN-CLAUDE  
DUCANCEL, FREDERIC  
APPLICANT: MENEZ, ANDRE  
TITLE OF INVENTION: MODIFIED BACTERIAL ALKALINE PHOSPHATASES  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/673,312  
FILING DATE: 28-JUN-1996  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: FR 95-07833  
FILING DATE: 29-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 846-380-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220

```
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 6162 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-673-312-9

Alignment Scores:
Pred. No.:      647      Length:      6162
Score:          43.50    Matches:      13
Percent Similarity: 46.34% Conservative: 6
Best Local Similarity: 31.71% Mismatches: 9
Query Match:    33.72% Indels:      13
                Gaps:      1
DB: 2

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-673-312-9 (1-6162)

Oy 1 SeraspAlaLeuGluaLaLeuAlaAspGlnThraSpAla----- 13
Db 1252 AATGACAGTGTACCAACCTGGCCACAGATGACCGACAAGCCATTGATTGTCAGTAA 1311
Oy 14 -----LeuGlnSerGlnGluAlaAlaValAlaValAlaAspAsnAla 27
Db 1312 AATGAGAAAGGCTTTTCCCTGCAGTGTGAAGGTCGTCATGCAATGAAACAGATCATGCT 1371
Oy 28 Ala 28
Db 1372 GCG 1374

RESULT 27
US-08-934-846-3
; Sequence 3, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
;   APPLICANT: Pearson, Stewart C.
;   APPLICANT: Greenwood, Rebecca C.
;   TITLE OF INVENTION: NOVEL folC
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Dechert Price & Rhoads
;   STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: US
;   ZIP: 19103
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FASTSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/934,846
;   FILING DATE:
;   CLASSIFICATION: 536
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Dickinson, Todd O
;   REGISTRATION NUMBER: 28,354
;   REFERENCE/DOCKET NUMBER: GM10088
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-994-2252
;   TELEFAX: 215-994-2222
;   TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1221 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
US-09-847-539a-6_COPY_59_86 (1-28) x US-09-238-557-3 (1-1221)
```

```
US-08-934-846-3

Alignment Scores:
Pred. No.:      104      Length:      1221
Score:          43.00    Matches:      5
Percent Similarity: 51.85% Conservative: 9
Best Local Similarity: 33.33% Mismatches: 13
Query Match:    33.33% Indels:      0
                Gaps:      0
DB: 2

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-934-846-3 (1-1221)

Oy 2 AspaLaLeuGluaLaLeuAlaAspGlnThraSpAlaLeuGlnSerGlnGluAlaAlaVal 21
Db 499 GACCATCAAGAAACCTTGGGTGATGCTGTAAGACATTGCAAGACAGAAAGCTGATATT 558
Oy 22 VallysAlaAspAsnAlaAla 28
Db 559 TTCAAGGCTGTGAAGAAGCA 579

RESULT 28
US-09-238-557-3
; Sequence 3, Application US/09238557
; Patent No. 6165472
; GENERAL INFORMATION:
;   APPLICANT: Pearson, Stewart C.
;   APPLICANT: Greenwood, Rebecca C.
;   TITLE OF INVENTION: NOVEL folC
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Dechert Price & Rhoads
;   STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: US
;   ZIP: 19103
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FASTSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/238,557
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/934,846
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Dickinson, Todd O
;   REGISTRATION NUMBER: 28,354
;   REFERENCE/DOCKET NUMBER: GM10088
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-994-2252
;   TELEFAX: 215-994-2222
;   TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1221 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
US-09-238-557-3

Alignment Scores:
Pred. No.:      104      Length:      1221
Score:          43.00    Matches:      9
Percent Similarity: 51.85% Conservative: 5
Best Local Similarity: 33.33% Mismatches: 13
Query Match:    33.33% Indels:      0
                Gaps:      0
DB: 4

US-09-847-539a-6_COPY_59_86 (1-28) x US-09-238-557-3 (1-1221)
```



```

1      PRIOR APPLICATION DATA:  US 08/081,448
2      APPLICATION NUMBER:  US 08/081,448
3      FILING DATE:  22-JUN-1993
4      ATTORNEY/AGENT INFORMATION:
5      NAME:  Highlander, Steven L.
6      REGISTRATION NUMBER:  37,642
7      REFERENCE/DOCKET NUMBER:  ARCD:090--1
8      TELECOMMUNICATION INFORMATION:
9      TELEPHONE:  (512) 418-5000
10     TELEFAX:  (512) 474-7577
11     INFORMATION FOR SEQ ID NO: 1:
12     SEQUENCE CHARACTERISTICS:
13     LENGTH: 1274 base pairs
14     TYPE: nucleic acid
15     STRANDEDNESS: single
16     TOPOLOGY: linear
17     FEATURE:
18     NAME/KEY:  CDS
19     LOCATION:  179..751
20     US-08-470-670A-1
21
22     Alignment Scores:
23     Pred. No.:
24     Score: 134
25     Percent Similarity: 42.50
26     Best Local Similarity: 65.52%
27     Query Match: 48.28%
28     DB: 32.95%
29     Gaps: 1
30
31     US-09-847-539A-6_COPY_59_86 (1-28) x US-08-470-670A-1 (1-1274)
32
33     Oy 2 AspaLaLeuguaLaLeuaLa-AspGlnThrAspaLa---LeugInserGluGuaLaLa 20
34         |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
35     Db 1230 GATTCACCTCCACGACGATGCGCAGCGGACGACGATCTCCATCTCCAGAAATGAAGAGTAGC 117
36
37     Oy 20 aValValLysAlaLaAspaSmaLaLa 28
38         ||||| |||:||||| |||
39     Db 1170 TGTTACCCCCACCAAGATCATCAGGCGC 1146
40
41     RESULT 33
42     US-08-461-511A-1/c
43     Sequence 1, Application US/08461511A
44     Patent No. 6303331
45     GENERAL INFORMATION:
46     APPLICANT: Thompson, Craig B.B.
47                 Boise, Lawrence H.
48     TITLE OF INVENTION: VERTEBRATE APOPTOSIS GENE: COMPOSITIONS
49                             AND METHODS
50     NUMBER OF SEQUENCES: 18
51     CORRESPONDENCE ADDRESS:
52     ADDRESSEE: Arnold, White & Durkee
53     STREET: P.O. Box 4433
54     City: Houston
55     STATE: Texas
56     COUNTRY: United States of America
57     Zip: 77210
58     COMPUTER READABLE FORM:
59     MEDIUM TYPE: Floppy disk
60     COMPUTER: IBM PC compatible
61     OPERATING SYSTEM: PC-DOS/MS-DOS
62     SOFTWARE: Patentin Release #1.0, Version #1.30
63     CURRENT APPLICATION DATA:
64     APPLICATION NUMBER: US/08/461,511A
65     FILING DATE: 05-Jun-1995
66     CLASSIFICATION: UNKNOWN
67     ATTORNEY/AGENT INFORMATION:
68     NAME: Highlander, Steven L.
69     REGISTRATION NUMBER: 37,642
70     REFERENCE/DOCKET NUMBER: ARCD:179
71     TELECOMMUNICATION INFORMATION:
72     TELEPHONE: (512) 418-3000
73     TELEFAX: (512) 474-7577
74     INFORMATION FOR SEQ ID NO: 1:

```

[illegible]

```

Alignment Scores:
Pred. No.: 134 Length: 1274
Score: 42.50 Matches: 14
Percent Similarity: 65.52% Conservative: 5
Best Local Similarity: 48.28% Mismatches: 8
Query Match: 32.95% Indels: 2
DB: 5 Gaps: 1

US-09-847-539A-6_COPY_59_86 (1-28) x PCT-US94-07089-1 (1-1274)

QY 2 AspaLaLeuglAaLLeuAlaAa-AspGlnrhrAspaLa---LeuGlnSrGjUGLAlAlAl 20
||||:|||| |:::||||| ||||| ::| |||||:::||||| ||
Db 1230 GATTCATCTCCACGGCATGGCAGCGCAGCATGCTCCATCTCCAGATGAATGAAGATGAC 1172
QY 20 aAaIvAllyAlAspaAspaAlaAla 28
|||| |:::| |||
Db 1170 TGTACCCCGACCATGATCAGAGGCGC 1146

RESULT 35
US-08-081-448-1/c
Sequence 1, Application US/08081448
Patent No. 5646008
GENERAL INFORMATION:
APPLICANT: Thompson, Craig B.
APPLICANT: Boise, Lawrence H.
TITLE OF INVENTION: Vertebrate Apoptosis Gene:
TITLE OF INVENTION: Compositions and Methods
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: 321 No. 5646008th Clark Street, Suite 800
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60610
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/081,448
FILING DATE: 19930622
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 5646008thrup, Thomas E.
REGISTRATION NUMBER: 33,268
REFERENCE/DOCKET NUMBER: ARCD090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-744-0090
TELEFAX: 312-755-4489
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 179..751
US-08-081-448-1

Alignment Scores:
Pred. No.: 138 Length: 1303
Score: 42.50 Matches: 14
Percent Similarity: 65.52% Conservative: 5
Best Local Similarity: 48.28% Mismatches: 8
Query Match: 32.95% Indels: 2
DB: 1 Gaps: 1

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-081-448-1 (1-1303)

```







REGISTRATION NUMBER: 32,323  
REFERENCE/DOCKET NUMBER: CM356M  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 513-627-2885  
TELEFAX: 513-627-0318  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1060 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 10..924  
US-08-081-328-1

Alignment Scores:  
Pred. No.: 130 Length: 1060  
Score: 42.00 Matches: 15  
Percent Similarity: 59.26% Conservative: 1  
Best local Similarity: 55.56% Mismatches: 9  
Query Match: 32.56% Indels: 2  
DB: 1 Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-081-328-1 (1-1060)

QY 2 AspaAlaLeuGluAlaLeuAlaAspGlnThrAspaAlaLeuGlnSerGluAlaAlaVal 21  
Db 522 GATGCCGCCGCTAGCGCGTGGCGGCGAGCGCGAAGTGGGAGT-----GCATCCGTC 469  
QY 22 ValLysAlaAspAsnAlaAla 28  
Db 468 GAAGATGCCGAGCGCCGCCGCC 448

Search completed: October 13, 2002, 06:32:14  
Job time : 1240.98 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 04:38:53 : Search time 7.63636 Seconds  
(without alignments)  
352.328 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALPALADQTDALQSEEAAYVKAADNA 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57.5	44.6	309	1 YAZON7	S-antigen precursor
2	57	44.2	412	2 G64685	hypothetical prote
3	57	44.2	412	2 B71831	hypothetical prote
4	51.5	39.9	213	2 JC1152	hypothetical prote
5	51	39.5	203	2 AH1080	hypothetical prote
6	51	39.5	203	2 AI1437	hypothetical prote
7	51	39.5	276	2 T29125	ketoreductase
8	51	39.5	677	2 S73798	MG260 homolog H91
9	51	39.5	4589	2 T14914	dynein beta heavy
10	50	38.8	617	2 G64972	yegA protein precu
11	50	38.8	617	2 G90987	suppressor of omp
12	50	38.8	617	2 B85833	suppressor of omp
13	50	38.8	1968	1 S05697	myosin heavy chain
14	49	38.0	142	2 AC0828	probable regulator
15	49	38.0	445	2 T06051	hypothetical prote
16	49	38.0	684	2 A84305	hypothetical prote
17	47.5	36.8	421	2 D95975	hypothetical prote
18	47	36.4	139	1 B70955	aspartate 1-decarb
19	47	36.4	192	2 T09236	nifz protein - Fra
20	47	36.4	192	2 JC4207	nifrogenase NifZ C
21	47	36.4	254	2 A71519	hypothetical prote
22	47	36.4	254	2 A81677	conserved hypochet
23	47	36.4	307	2 AB2807	conserved hypochet
24	47	36.4	333	2 A97586	hypothetical prote
25	47	36.4	1013	2 G87236	conserved integral
26	47	36.4	2429	1 SJHXA	spectrin alpha cha
27	46.5	36.0	536	2 D84325	Htr17 transducer (
28	46.5	36.0	784	2 T51759	glutamate--ammoni
29	46	35.7	106	2 A81203	conserved hypochet

30	46	35.7	123	2 E90066	aspartate 1-decarb
31	46	35.7	345	2 A87375	hypothetical prote
32	46	35.7	353	2 A72410	chorismate mutase/
33	46	35.7	425	2 T25873	hypothetical prote
34	46	35.7	526	2 C83488	probable semialden
35	46	35.7	539	2 G83720	nickel transport s
36	46	35.7	618	2 AE0770	probable outer mem
37	46	35.7	732	2 F84394	helicase [imported
38	46	35.7	764	2 A84328	Htr2 transducer (1
39	46	35.7	765	1 T44846	transducer protein
40	46	35.7	787	2 S73873	probable lipoprote
41	45.5	35.3	480	2 F81382	pyruvate kinase (E
42	45.5	35.3	557	2 A83604	transport ATP-bind
43	45.5	35.3	572	2 T00707	transcription init
44	45.5	35.3	2186	2 H89960	hypothetical prote
45	45	34.9	84	2 E84345	hypothetical prote

## ALIGNMENTS

## RESULT 1

YAZON7  
S-antigen precursor - malaria parasite (Plasmodium falciparum) (strain NF7/Ghana)  
C:Species: Plasmodium falciparum  
C>Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 09-Jun-2000  
C/Accession: B22011  
R/Comment: A.F.: Saint, R.B.: Coppel, R.L.: Brown, G.V.: Anders, R.F.: Kemp, D.J.  
Cell 40, 775-783, 1985  
A>Title: Conserved sequences flank variable tandem repeats in two S-antigen genes of P.  
A/Reference number: A90863; MID:85176931  
A/Accession: B22011  
A/Molecule type: DNA  
A/Residues: 1-309 <COM>  
A/Cross-references: GB:M10130; NID:g160670; PIDN:AAA29758.1; PID:g160671  
A/Experimental source: clone NF7.S  
A/Note: The intact NF7 S-antigen contains about 35 more of the 8-residue repeats  
C/Comment: The S-antigen is secreted by the parasite.  
C/Superfamily: Plasmodium S-antigen  
C/Keywords: malaria; surface antigen; tandem repeat  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-309/Product: S-antigen #status predicted <MAT>  
F:27-286/Region: 8-residue repeats (S-D-E-A-E-A-L/R-K)  
F:257-271,272-286/Region: 15-residue repeats

Query Match 44.6%; Score 57.5; DB 1; Length 309;  
Best local similarity 50.0%; Pred. No. 2.1;  
Matches 14; Conservative 6; Mismatches 7; Indels 1; Gaps 1;

QY 1 SDALPALADQTDALQSEEAAYVKAADNA 27  
DB 97 SDEAALKSDAEALKSDEAEARKSDEA 124

## RESULT 2

G64685  
hypothetical protein HP1327 - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
C/Accession: G64685  
R/Comment: S.: Loftus, B.: Richardson, D.; Dodson, R.; Khailak, H.G.; Glodok, A.; McKee  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Matthey,  
Nature 388, 539-547, 1997  
A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,  
A>Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A/Reference number: A64520; MID:97334467  
A/Accession: G64685  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-412 <COM>  
A/Cross-references: GB:AE000634; GB:AE000511; NID:g2314489; PIDN:AAD08377.1; PID:g231

```
Query Match 44.2%; Score 57; DB 2; Length 412;
Best Local Similarity 52.2%; Pred. No. 3.3;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 4 LEALADOTDALQSEEAAYVKADN 26
   ::||:| ||| ||| | | | |
Db 49 IOALQEQIDALDSQEKVSKMDN 71

RESULT 3
hypotheical protein jhp1247 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: Strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: B71831
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.R.;
: Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557
A:Accession: B71831
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <ARN>
A:Cross-references: GB:AE001548; GB:AE00439; NID:g4155845; PIDN:AAD06820.1; PID:g415584
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp1247

Query Match 44.2%; Score 57; DB 2; Length 412;
Best Local Similarity 52.2%; Pred. No. 3.3;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 4 LEALADOTDALQSEEAAYVKADN 26
   ::||:| ||| ||| | | | |
Db 49 IOALQEQIDALDSQEKVSKMDN 71

RESULT 4
JCI152
hypotheical protein, 23.6K - Agrobacterium tumefaciens insertion sequence IS1131
C:Species: Agrobacterium tumefaciens
C:Date: 05-Mar-1993 #sequence_revision 05-Mar-1993 #text_change 17-Jul-1998
C:Accession: JCI152
R:Wadiko, H.
Gene 114, 229-233, 1992
A:Title: Sequence analysis of an insertion element, IS1131, isolated from the nopaline-t
A:Reference number: JCI150; MUID:92290280
A:Accession: JCI152
A:Molecule type: DNA
A:Residues: 1-213 <MAB>
A:Cross-references: GB:M82868
A:Experimental source: strain P022; plasmid T1; insertion sequence IS1131
C:Genetics:
A:Mobile element: insertion sequence IS1131

Query Match 39.9%; Score 51.5; DB 2; Length 213;
Best Local Similarity 52.0%; Pred. No. 9.3;
Matches 13; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

OY 3 ALEAL-ADOTDALQSEEAAYVKADN 26
   || || | | | | | | | |
Db 14 ALRALVAEQAKLESQEAIVIKRDS 38

RESULT 5
AHI080
hypotheical protein lmo0047 [imported] - Listeria monocytogenes (strain ECD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AHI080
R:Glasier, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
```

```
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maltournam, A.;
Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AHI080
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <GLA>
A:Cross-references: GB:NC_003210; PIDN:GAC98262.1; PID:g16409406; GSPDB:GN00177
A:Experimental source: strain ECD-e
C:Genetics:
A:Gene: lmo0047

Query Match 39.5%; Score 51; DB 2; Length 203;
Best Local Similarity 38.5%; Pred. No. 10;
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

OY 1 SDALEADOTDALQSEEAAYVKADN 26
   :: ||| ||| | | | | |
Db 109 AETKEQLSDERTDRDREDAGVEXEN 134

RESULT 6
AII437
hypotheical protein lln0040 [imported] - Listeria innocua (strain C1p11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AII437
R:Glasier, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maltournam, A.;
Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AII437
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC95273.1; PID:g16412461; GSPDB:GN00178
A:Experimental source: strain C1p11262
C:Genetics:
A:Gene: lln0040

Query Match 39.5%; Score 51; DB 2; Length 203;
Best Local Similarity 38.5%; Pred. No. 10;
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

OY 1 SDALEADOTDALQSEEAAYVKADN 26
   :: ||| ||| | | | | |
Db 109 AETKELEADETDKLKEEDGVEXEN 134

RESULT 7
T29125
ketocyl reductase homolog - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 20-Jun-2000
C:Accession: T29125
R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z17215
A:Accession: T29125
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-276 <PAR>
A:Cross-references: EMBL:AL031350; PIDN:CAA20507.1
C:Genetics:
A:Note: SC1P2.16c
```

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 39.5%; Score 51; DB 2; Length 276;  
Best Local Similarity 48.0%; Pred. No. 14;  
Matches 12; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

OY 1 SDALADQTDALQSEAAVVRAD 25  
:|||||:|:|:|:|  
Db 40 ADALHADADELRAHRSVRVDYIAD 64

#### RESULT 8

MG260 homolog H91-ori677 - Mycoplasma pneumoniae (strain ATCC 29342)

C:Species: Mycoplasma pneumoniae

A:Variety: ATCC 29342

C:Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999

C:Accession: S73798

R:Himmelreich, R.; Hilbert, H.; Piagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.

Nucleic Acids Res. 24, 4420-4449, 1996

A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S73797; MUID:97105885

A:Accession: S73798

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Residues: 1-677 <HMM>

A:Cross-references: EMBL:AE000047; GB:U00089; NID:91674162; PIDN:AB96120.1; PID:9167416

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996

C:Genetics:

A:Genetic code: SGC3

C:Superfamily: hypothetical protein MG185

Query Match 39.5%; Score 51; DB 2; Length 677;  
Best Local Similarity 50.0%; Pred. No. 37;  
Matches 11; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 3 ALALADQTDALQSEAAVVRAD 24  
:|:|:|:|:|:|:|:|:|:|  
Db 486 ATEVLEQDSNTQSEAAVVRAD 507

#### RESULT 9

T14914  
dynein beta heavy chain - Tetrahymena thermophila

C:Species: Tetrahymena thermophila

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000

C:Accession: T14914

R:Lincoln, L.M.; Gibson, T.M.; Asai, D.J.; Forney, J.D.

submitted to the EMBL Data Library, June 1999

A:Description: A gene knockout reveals that dynein beta heavy chain is required in Tetra

A:Reference number: Z18264

A:Accession: T14914

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-4589 <LIN>

A:Cross-references: EMBL:AF072878; NID:95209335; PID:93309593; PIDN:AA226117.1

C:Genetics:

A:Gene: DYH4

A:Genetic code: SGC5

A:introns: 286/3; 666/3; 4360/2; 4535/3

A:Superfamily: dynein heavy chain, ciliary

C:Keywords: P-loop

Query Match 39.5%; Score 51; DB 2; Length 4589;  
Best Local Similarity 60.0%; Pred. No. 2.8e+02;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 DALEADQTDALQSEAAV 21  
:|:|:|:|:|:|:|:|:|:|  
Db 3205 EATLEALPALRSRAAV 3224

#### RESULT 10

G64972

yegA protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 21-Jul-2000

C:Accession: G64972; B42940; S77641

R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A:Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: G64972

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-617 <BLAT>

A:Cross-references: GB:AE000296; GB:U00096; NID:91788373; PIDN:AACT5125.1; PID:917883

A:Experimental source: strain K-12, substrain MG1655

R:Mang, L.; Weiss, B.

J. Bacteriol. 174, 5647-5653, 1992

A:Title: dcd (dcp deaminase) gene of Escherichia coli: mapping, cloning, sequencing,

A:Reference number: A42940; MUID:92380941

A:Accession: B42940

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-32 <MAN>

A:Note: sequence extracted from NCBI backbone (NCBIN:112699, NCBI:112721)

R:Mista, R.; Miao, Y.

Mol. Microbiol. 16, 779-788, 1995

A:Title: Molecular analysis of asmA, a locus identified as the suppressor of OmpF ass

A:Reference number: S77641; MUID:96065706

A:Accession: S77641

A:Molecule type: DNA

A:Residues: 1-353; 'NEFYRCRANQVGFAGNYSTAGOT' <NIS>

A:Cross-references: EMBL:U11035; NID:9684919; PIDN:AD14778.1; PID:94262670

A:Experimental source: strain K-12

C:Genetics:

A:Gene: yegA; asmA

C:Function:

A:Description: involved in the assembly of outer membrane proteins

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-380/Product: yegA protein #status predicted <MAT>

Query Match 38.8%; Score 50; DB 2; Length 617;  
Best Local Similarity 45.8%; Pred. No. 46;  
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 3 ALALADQTDALQSEAAVVRAD 26  
:|:|:|:|:|:|:|:|:|:|  
Db 109 AVIDLTPQTEAVRSDDAVAPRDN 132

#### RESULT 11

G90987  
suppressor of ompF assembly mutants [imported] - Escherichia coli (strain O157:H7, su

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: G90987

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C

gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: G90987

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-617 <HAV>

A:Cross-references: GB:BA000007; PIDN:BA36294.1; PID:913362340; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RMd 050952

C:Genetics:

A:Gene: ECs2871

Query Match 38.8%; Score 50; DB 2; Length 617;  
Best Local Similarity 45.8%; Pred. No. 46;  
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;



A:Map position: 4  
A:Introns: 37/3; 270/1; 312/2  
C:Superfamily: Arabidopsis thaliana hypothetical protein F19H22.10

Query Match  
Best Local Similarity 43.5%; Score 49; DB 2; Length 445;  
Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAAYKADNA 27  
||||| : : : : :  
Db 78 EALASRDEALEQRDKALSERDNA 100

## RESULT 16

A:Map position: 4  
A:Introns: 37/3; 270/1; 312/2  
C:Superfamily: Arabidopsis thaliana hypothetical protein F19H22.10  
Query Match  
Best Local Similarity 43.5%; Score 49; DB 2; Length 445;  
Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 2 DALEALADQTDAL--QSEEAAYKADNA 27  
||||| : : : : :  
Db 604 DALHRRADSTDVLLGAGDALRVSRSDA 631

## RESULT 17

A:Map position: 4  
A:Introns: 37/3; 270/1; 312/2  
C:Superfamily: Arabidopsis thaliana hypothetical protein F19H22.10  
Query Match  
Best Local Similarity 43.5%; Score 49; DB 2; Length 445;  
Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Query Match  
Best Local Similarity 43.5%; Score 49; DB 2; Length 445;  
Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Best Local Similarity 53.8%; Pred. No. 67;  
Matches 14; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

QY 4 LEALADQTDALQSEEAAYKADNNA 28  
||||| : : : : :  
Db 240 LDALADRLSLQSEEAALAKRAETQA 265

## RESULT 18

A:Map position: 4  
A:Introns: 37/3; 270/1; 312/2  
C:Superfamily: Arabidopsis thaliana hypothetical protein F19H22.10  
Query Match  
Best Local Similarity 43.5%; Score 49; DB 2; Length 445;  
Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 7 LADQTDALQSEEAAYKADNNA 28  
||||| : : : : :  
Db 32 LMDADADLEGEQVTIVDIDNGA 53

## RESULT 19

A:Map position: 4  
A:Introns: 37/3; 270/1; 312/2  
C:Superfamily: Arabidopsis thaliana hypothetical protein F19H22.10  
Query Match  
Best Local Similarity 43.5%; Score 49; DB 2; Length 445;  
Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 4 LEALADQTDALQSEEAAYKADNNA 28  
||||| : : : : :  
Db 74 LEAEDADBDQAQEPAAATPTAATAA 98

RESULT 20  
JC4207  
nltrogenase nlfz chain - Frankia sp.  
N:Alternate names: Nlfz protein  
C:Species: Frankia sp.  
C:Date: 10-Sep-1995 #sequence\_revision 27-Oct-1995 #text\_change 22-Oct-1999  
C:Accession: JC4207  
R:Harriot, O.T.; Hosted, T.J.; Benson, D.R.  
Gene 161, 63-67, 1995  
A:Title: Sequences of nlfz, nlfw, nlfz, nlfB and two ORF in the Frankia nitrogen fixative  
A:Reference numbers: JC4203; MUID:95369734  
A:Accession: JC4207  
A:Molecule type: DNA  
A:Residues: 1-192 <NAR>  
A:Cross-references: GB:L29299; NID:g497430; PIDN:AAC82974.1; PID:g497435  
A:Comment: This protein has an Ala and Pro rich region at its carboxyl-terminal region.  
C:Genetics:  
A:Gene: nlfz  
C:Keywords: nitrogen fixation

Query Match 36.4%; Score 47; DB 2; Length 192;  
Best Local Similarity 52.0%; Pred. No. 35;  
Matches 13; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 4 LEALADOTDALQSEAAVVRADNA 28  
Db 74 LEAAEDADATQAPPAATPTAAAAA 98

RESULT 21  
A71519  
hypothetical protein CT398 - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C:Accession: A71519  
R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac  
A:Reference number: A71570; MUID:99000809  
A:Accession: A71519  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-254 <ARN>  
A:Cross-references: GB:AE001333; GB:AE001273; NID:g3328823; PIDN:AAC67995.1; PID:g332882  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: CT398

Query Match 36.4%; Score 47; DB 2; Length 254;  
Best Local Similarity 36.4%; Pred. No. 46;  
Matches 8; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

OY 4 LEALADOTDALQSEAAVVRAD 25  
Db 67 IQEISDQINKLENOQAARVKKMD 88

RESULT 22  
A81677  
conserved hypothetical protein TC0677 [Imported] - Chlamydia muridarum (strain N199)  
C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: A81677  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AK39.  
A:Reference number: A81500; MUID:20150255  
A:Accession: A81677  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-254 <TET>

A:Cross-references: GB:AE002336; GB:AE002160; NID:g7190707; PIDN:AAE39497.1; PID:g719  
A:Experimental source: strain N199 (Mopn)  
C:Genetics:  
A:Gene: TC0677

Query Match 36.4%; Score 47; DB 2; Length 254;  
Best Local Similarity 36.4%; Pred. No. 46;  
Matches 8; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

OY 4 LEALADOTDALQSEAAVVRAD 25  
Db 67 IQEISDQINKLENOQAARVKKMD 88

RESULT 23  
AB2807  
conserved hypothetical protein Atu1876 [Imported] - Agrobacterium tumefaciens (strain  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AB2807  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; Mccl  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AB2807  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-307 <KUR>  
A:Cross-references: GB:AE008688; PIDN:AAL42872.1; PID:g17740324; GSPDB:GNO0186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu1876  
A:Map position: circular chromosome

Query Match 36.4%; Score 47; DB 2; Length 307;  
Best Local Similarity 47.8%; Pred. No. 57;  
Matches 11; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 2 DALEALADOTDALQSEAAVVRKA 24  
Db 79 DQLEALADOTDALQSEAAVVRKA 101

RESULT 24  
A97586  
hypothetical protein AGR\_C\_3443 [Imported] - Agrobacterium tumefaciens (strain C58, C  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: A97586  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Gold  
A.; Liu, F.; Mollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Leppas, C.; Markelz,  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
A:Reference number: A97359; PMID:11743194  
A:Accession: A97586  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-333 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK87642.1; PID:g15156994; GSPDB:GNO0169  
C:Genetics:  
A:Gene: AGR\_C\_3443  
A:Map position: circular chromosome

Query Match 36.4%; Score 47; DB 2; Length 333;  
Best Local Similarity 47.8%; Pred. No. 62;  
Matches 11; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 2 DALEALADOTDALQSEAAVVRKA 24  
Db 79 DQLEALADOTDALQSEAAVVRKA 101



Db 105 D0LEALQOTDAMKVRADIAFA 127

RESULT 25

conserved integral membrane protein M2617 (imported) - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: G87236

R:Colet, S.T.; Eismeleier, K.; Parkhill, J.; James, K.D.; Thomson, M.R.; Wheeler, P.R.; Ho R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Fellwell, T.; Fraser, A.; Hamlin, N.; Holroyd, eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sc A:Title: Massive gene decay in the leprosy bacillus.

A:Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: G87236

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1013 <STO>

A:Cross-references: GB:AL450380; MID:g13093843; PIDN:CAC32149.1; GSPDB:GN00147 C:Genetics: M2617

Query Match 36.4%; Score 47; DB 2; Length 1013;

Best Local Similarity 62.5%; Pred. No. 2e+02;

Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 7 LADQTDALQSEAAV 22

Db 931 LADWDAIEPTAAV 946

RESULT 26

SJHUA

Spectrin alpha chain - human

C:Species: Homo sapiens (man)

C:Date: 17-May-1985 #sequence\_revision 30-Jun-1992 #text\_change 16-Jul-1999

C:Accession: A35716; A45755; A23533; A91528; A42872; A02965; S13138; A05282; A23965; A27 R:Saht, K.E.; Laurila, P.; Kotila, L.; Scarpa, A.L.; Coupal, E.; Leto, T.L.; Linnenbach, J. Biol. Chem. 265, 4434-4443, 1990

A:Title: The complete cDNA and polypeptide sequences of human erythroid alpha-spectrin.

A:Reference number: A35716; MUID:90170949

A:Accession: A35716

A:Molecule type: mRNA

A:Residues: 1-2429 <SAH>

A:Cross-references: GB:J05244

R:Saht, K.E.; Tope, T.; Scarpa, A.; Laughinghouse, K.; Marchesi, S.L.; Agre, P.; Linne J. Clin. Invest. 84, 1243-1252, 1989

A:Title: Sequence and exon-intron organization of the DNA encoding the alpha domain of A:Reference number: A45755; MUID:90003318

A:Accession: A45755

A:Molecule type: DNA

A:Residues: 1-394; G', 396-533 <SAZ>

A:Cross-references: GB:M29983

A:Note: the authors translated the codon GGT for residue 395 as Ala

R:Linnebach, A.U.; Speicher, D.W.; Marchesi, V.T.; Forget, B.G.

Proc. Natl. Acad. Sci. U.S.A. 83, 2397-2401, 1986

A:Title: Cloning of a portion of the chromosomal gene for human erythrocyte alpha-spectr A:Reference number: A23533; MUID:86205962

A:Accession: A23533

A:Molecule type: DNA

A:Residues: 320-450 <LIN>

A:Cross-references: GB:M13233; MID:g182242; PIDN:AAA53103.1; PID:g182243

R:Curtis, P.J.; Palumbo, A.; Ming, J.; Fraser, P.; Cloe, L.; Meo, P.; Shane, S.; Rovera, Gene 36, 357-362, 1985

A:Title: Sequence comparison of human and murine erythrocyte alpha-spectrin cDNA.

A:Accession: A91528

A:Reference number: A91528; MUID:86083178

A:Molecule type: mRNA

A:Residues: 1451-1687 <CUR>

A:Cross-references: GB:M1049; MID:g338310; PIDN:AAA60569.1; PID:g553648

R:Speicher, D.W.; Weglartz, L.; Desilva, T.M.

J. Biol. Chem. 267, 14775-14782, 1992

A:Title: Properties of human red cell spectrin heterodimer (side-to-side) assembly an A:Reference number: A42872; MUID:92340516

A:Accession: A42872

A:Molecule type: protein

A:Residues: 7-16; 46-55; 680-689; 1047-1056; 1921-1930 <SPI>

R:Speicher, D.W.; Davis, G.; Marchesi, V.T.

J. Biol. Chem. 258, 14938-14947, 1983

A:Title: Structure of human erythrocyte spectrin. II. The sequence of the alpha-I dom A:Reference number: A92408; MUID:8408788

A:Accession: A02965

A:Molecule type: protein

A:Residues: 7-601 <SPC>

R:Speicher, D.W.; Davis, G.; Vurchenco, P.D.; Marchesi, V.T.

J. Biol. Chem. 258, 14931-14937, 1983

A:Title: Structure of human erythrocyte spectrin. I. Isolation of the alpha-I domain A:Reference number: S13138; MUID:8408788

A:Accession: S13138

A:Molecule type: protein

A:Residues: 7-92; 'X', 94-96, 110-151; 317-342; 345-366, 'Z', 368-370, 'X', 372-373; 387-434; 45 R:Lusitani, D.M.; Qataishat, N.; Labraque, C.C.; Yu, R.N.; Davis, J.; Kelley, M.R.; Fun J. Biol. Chem. 269, 25955-25958, 1994

A:Title: The first human alpha-spectrin structural domain begins with serine.

A:Reference number: A38928; MUID:95014412

A:Contents: annotation; determination of structural domain

C:Comment: Spectrin associates with band 4.1 and actin to form the cytoskeletal super S of approximately 106 residues each.

C:Genetics:

A:Gene: GDB:SPFAL

A:Cross-references: GDB:119601; OMIM:182860

A:Map position: 1q21-1q21

C:Superfamily: spectrin alpha chain; calmodulin repeat homology; SH3 homology; spectr C:Keywords: actin binding; cytoskeleton; duplication; EF hand; erythrocyte; heterodim F:52-156/Domain: spectrin/dystrophin repeat homology <SP1>

F:157-262/Domain: spectrin/dystrophin repeat homology <SP2>

F:263-368/Domain: spectrin/dystrophin repeat homology <SP3>

F:369-474/Domain: spectrin/dystrophin repeat homology <SP4>

F:475-580/Domain: spectrin/dystrophin repeat homology <SP5>

F:581-685/Domain: spectrin/dystrophin repeat homology <SP6>

F:686-791/Domain: spectrin/dystrophin repeat homology <SP7>

F:792-897/Domain: spectrin/dystrophin repeat homology <SP8>

F:898-993/Domain: spectrin/dystrophin repeat homology #status atypical <SP9>

F:994-1031/Domain: SH3 homology <SH3>

F:1081-1181/Domain: spectrin/dystrophin repeat homology <SP10>

F:1182-1287/Domain: spectrin/dystrophin repeat homology <SP11>

F:1288-1393/Domain: spectrin/dystrophin repeat homology <SP12>

F:1394-1498/Domain: spectrin/dystrophin repeat homology <SP13>

F:1499-1604/Domain: spectrin/dystrophin repeat homology <SP14>

F:1605-1710/Domain: spectrin/dystrophin repeat homology <SP15>

F:1711-1816/Domain: spectrin/dystrophin repeat homology <SP16>

F:1817-1925/Domain: spectrin/dystrophin repeat homology <SP17>

F:1926-2032/Domain: spectrin/dystrophin repeat homology <SP18>

F:2041-2146/Domain: spectrin/dystrophin repeat homology <SP19>

F:2155-2257/Domain: spectrin/dystrophin repeat homology <SP20>

F:2270-2302/Domain: calmodulin repeat homology <EF1>

F:2313-2345/Domain: calmodulin repeat homology <EF2>

Query Match 36.4%; Score 47; DB 1; Length 2429;

Best Local Similarity 40.9%; Pred. No. 5e+02;

Matches 9; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Oy 2 DALEALQOTDAMKVRADIAFA 23

Db 953 DSMKALRQANACOOQAAVPE 974

RESULT 27

DB4325

Htr17 transducer (imported) - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001

C:Accession: DB4325

R:Ng, W.V.; Kennedy, S.P.; Mahatras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky

; Leithauer, B.; Koller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo  
Jung, K.H.; Alam, M.; Freltas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of *Halobacterium* species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: D84325  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-536 <STO>  
A:Cross-references: GB:AE004437; NID:g10581193; PIDN:AA619968.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: htr17  
C:Superfamily: Halobacterium salinarum transducer protein htr1

Query Match 36.0%; Score 46.5; DB 2; Length 536;  
Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
Matches 13; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

OY 2 DALEALADQTDALQSEEAAYVAKDNA 27  
DB 295 DAALAAVTQMTD-IQSETEAVVSEANA 319

RESULT 28  
T51759  
[glutamate--ammonia-lyase] adenylyltransferase (EC 2.7.7.42) [Imported] - Streptomyces  
C:Species: Streptomyces coelicolor  
C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
C:Accession: T51759  
R:Link, D.; Falke, D.; Wohlleben, W.; Engels, A.  
Microbiology 145, 2313-2322, 1999  
A:Title: Nitrogen metabolism in Streptomyces coelicolor A3(2): modification of glutamine  
A:Reference number: Z25448  
A:Accession: T51759  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-784 <PIN>  
A:Cross-references: EMBL:Y17736; PIDN:CAA76840.2  
C:Genetics:  
A:Gene: gltE  
C:Keywords: nucleotidyltransferase

Query Match 36.0%; Score 46.5; DB 2; Length 784;  
Best Local Similarity 56.5%; Pred. No. 1.8e+02;  
Matches 13; Conservative 2; Mismatches 3; Indels 5; Gaps 1;

OY 2 DALEALA-----DQTDALQSEEA 19  
DB 391 DALEALACGYGRTDAALQDEA 413

RESULT 29  
A81203  
conserved hypothetical protein NMB0404 [Imported] - *Neisseria meningitidis* (strain MC58  
C:Species: *Neisseria meningitidis*  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: A81203  
R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizzi, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.; Ve  
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755  
A:Accession: A81203  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-106 <TEP>  
A:Cross-references: GB:AE002396; GB:AE002098; NID:g7225622; PIDN:AAF40843.1; PID:g722562  
C:Genetics:  
A:Experimental source: serogroup B, strain MC58  
A:Gene: NMB0404

Query Match 35.7%; Score 46; DB 2; Length 106;  
Best Local Similarity 42.3%; Pred. No. 25;  
Matches 11; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

OY 3 ALEALADQTDALQSEEAAYVAKDNA 28  
DB 65 ALFARLEKLEMAQNPERALAEAEAA 90

RESULT 30  
E90066  
aspartate 1-decarboxylase [Imported] - *Staphylococcus aureus* (strain N315)  
C:Species: *Staphylococcus aureus*  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: E90066  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O  
ma, A.; Mizutani-Uli, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: E90066  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-123 <KUR>  
A:Cross-references: GB:BA000018; PID:g13702554; PIDN:BA843695.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: pand  
C:Superfamily: aspartate 1-decarboxylase

Query Match 35.7%; Score 46; DB 2; Length 123;  
Best Local Similarity 42.9%; Pred. No. 30;  
Matches 12; Conservative 5; Mismatches 7; Indels 4; Gaps 1;

OY 1 SDALFALADQTDALQSEEAAYVAKDNA 28  
DB 26 SDILEAV----DILPEKVAIVNNNGA 49

RESULT 31  
A87375  
hypothetical protein CC1013 [Imported] - *Caulobacter crescentus*  
C:Species: *Caulobacter crescentus*  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: A87375  
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete genome sequence of *Caulobacter crescentus*.  
A:Reference number: A87249; MUID:21173698; PMID:1125647  
A:Accession: A87375  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-345 <STO>  
A:Cross-references: GB:AE005673; NID:g13422303; PIDN:AAK22997.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC1013

Query Match 35.7%; Score 46; DB 2; Length 345;  
Best Local Similarity 43.5%; Pred. No. 88;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 6 ALADQTDALQSEEAAYVAKDNA 28  
DB 303 ALAEQFDRLHADKAAPAKMGTA 325

RESULT 32  
A72410  
chorismate mutase/prephenate dehydratase - *Thermotoga maritima* (strain MSB8)





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:48:12 : Search time 39.5294 Seconds

(without alignments)  
1216.148 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDLEALADQTDALQSEEAAYKADNNA 28

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL-frame+ p2n.model -DEV-xlh  
-O/cgnt2\_1/USPRO.spool/US09847539/runtat\_10102002\_093105\_4982/app-query.fasta.1.526  
-DB-N-geneseq\_032802 -QPMF-fastlap -SUFFIX-std.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS-bits -START=1 -END=-1 -MATRIX-blosum62 -TRANS-humann40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIC=40  
-MODE-LOCAL -OUTPMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USR-US09847539.ecgn\_1.1\_48\_etunat\_10102002\_093105\_4982 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAMP -LARGEDUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV-TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7  
-YCAPOP=10 -YCAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_032802: \*  
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT: \*  
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT: \*  
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT: \*  
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT: \*  
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT: \*  
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT: \*  
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT: \*  
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT: \*  
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT: \*  
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT: \*  
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT: \*  
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT: \*  
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT: \*  
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT: \*  
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT: \*  
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT: \*  
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT: \*  
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT: \*  
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT: \*  
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT: \*  
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: \*  
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: \*  
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT: \*  
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	129	100.0	469	21	AAD00562	Streptococcus pyog
2	129	100.0	504	21	AAD00564	Streptococcus pyog
3	129	100.0	654	21	AAD00560	Streptococcus pyog
4	129	100.0	764	21	AAD00559	Streptococcus pyog
5	129	100.0	777	21	AAD00561	Streptococcus pyog
6	129	100.0	853	21	AAD00563	Streptococcus pyog
7	57	44.2	576	18	AA767618	H. pylori secreted
8	57	44.2	666	18	AA767891	H. pylori secreted
9	57	44.2	1239	19	AAK30461	H. pylori secreted
10	54	41.9	3349	23	ABL13527	Drosophila melanog
11	54	41.9	6398	23	ABL13526	Drosophila melanog
12	52	40.3	1002	21	AAK54228	Arabidopsis thalia
13	52	40.3	1230	21	AAK33266	Arabidopsis thalia
14	52	40.3	16135	23	ABL25211	Drosophila melanog
15	52	40.3	18603	23	ABL25210	Drosophila melanog
16	51.5	39.9	25461	23	ABL01876	Drosophila melanog
17	51	39.5	1960	23	ABL12683	Drosophila melanog
18	51	39.5	2618	23	ABL27360	Drosophila melanog
19	51	39.5	4221	23	ABL12682	Drosophila melanog
20	51	39.5	2944528	24	ABM03041	Listeria monocytog
21	50	38.8	3711	23	AAK94455	DNA encoding novel
22	48	37.2	1616	23	ABL25849	Drosophila melanog
23	48	37.2	4229	23	ABL25848	Drosophila melanog
24	47	36.4	4403765	22	AA199683	Mycobacterium tube
25	47	36.4	4411529	22	AA199682	Mycobacterium tube
26	46	35.7	636	21	AAK1343	Aspergillus niger
27	46	35.7	2115	22	AAK65320	C glutamicum codin
28	46	35.7	2160	22	AAK65521	C glutamicum codin
29	46	35.7	4549	18	AAK74485	Staphylococcus aur
30	46	35.7	69936	21	AAK81479	N. meningitidis pa
31	46	35.7	349980	21	AAK21607	Neisseria meningit
32	46	35.7	349980	22	AAK68525	C glutamicum codin
33	46	35.7	1437668	21	AAK81490	N. meningitidis B
34	45.5	35.3	234	22	ABA74678	Human foetal liver
35	45.5	35.3	234	22	AAK23154	Human brain expres
36	45.5	35.3	234	22	AAK49325	Human bone marrow
37	45.5	35.3	234	22	AA155172	Human bone marrow
38	45.5	35.3	544	22	ABA62184	Human foetal liver
39	45.5	35.3	544	22	AAK10508	Human brain expres
40	45.5	35.3	544	22	AAK36405	Human bone marrow
41	45.5	35.3	544	22	AA142139	Probe #10825 used
42	45.5	35.3	1719	21	AAK43055	Arabidopsis thalia
43	45.5	35.3	6228	23	AAK52178	Staphylococcus aur
44	45.5	35.3	6561	23	AAK55178	Staphylococcus aur
45	45.5	35.3	8009	18	AAK74372	Staphylococcus aur

# ALIGNMENTS

RESULT 1  
AAD00562  
ID AAD00562 standard: DNA: 469 BP.  
XX  
AC AAD00562:  
XX  
DT 29-AUG-2000 (first entry)  
XX  
XX Streptococcus pyogenes strain API partial GRAB protein encoding DNA.  
DE Streptococcus pyogenes strain API partial GRAB protein encoding DNA.  
XX  
XX GRAB protein: protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection; ds.  
XX  
XX Streptococcus pyogenes.  
OS  
XX Key Location/Qualifiers  
FH CDS 1..468  
FT

```
FT FT /*tag= a
FT FT /product= "GRAB protein"
FT FT /partial
XX XX
PN WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999: 99WO-GB03631.
XX
XX 02-NOV-1998: 98GB-0023975.
XX
XX (ACT1-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI: 2000-365572/31.
XX
XX P-PSDB: AAY71044.
XX
XX New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein -
XX
XX Claim 13; Page 64; 67pp: English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A streptococcus. Antibodies against GRAB
XX are useful for treating Streptococcus pyogenes infections. The protein
XX is also useful for purifying alpha2M from a sample. The present sequence
XX is a DNA encoding partial GRAB protein from S. pyogenes strain A1.
XX
XX The protein has alpha2M binding region and is useful in vaccine
XX composition.
XX
SQ Sequence 469 BP; 180 A; 90 C; 99 G; 100 T; 0 other;

Alignment Scores:
Pred. No.: 4,46e-13 Length: 469
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AAD00562 (1-469)
QY 1 SerASPAlaLeuGluAlaLeuAlaAspGlnThrASPAlaLeuGlnSerGluGluAlaAla 20
DB 172 TCAGATGCGCTTGAAGCATTAGCGGATCAAGACAGACGTTTACAAATCAGAGAAAGCTGCG 231
QY 21 ValValIysAlaIAspAsnAlaAla 28
DB 232 GTTGTTAAGCGGATTAACGCTGCT 255

RESULT 2
AAD00564
ID AAD00564 standard; DNA: 504 BP.
XX
XX AAD00564;
XX
XX 29-AUG-2000 (first entry)
XX
XX Streptococcus pyogenes strain KTL3 partial GRAB protein encoding DNA.
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
XX alpha2-macroglobulin; group A Streptococcus; GAS; antidiabetic;
XX immune response; Streptococcus pyogenes infection; ds.
XX
XX Streptococcus pyogenes.
XX
XX Key Location/Qualifiers
FH
```

```
FT FT CDS 1..504
FT FT /*tag= a
FT FT /product= "GRAB protein"
FT FT /partial
XX XX
PN WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999: 99WO-GB03631.
XX
XX 02-NOV-1998: 98GB-0023975.
XX
XX (ACT1-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI: 2000-365572/31.
XX
XX P-PSDB: AAY71046.
XX
XX New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein -
XX
XX Claim 13; Page 65; 67pp: English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A streptococcus. Antibodies against GRAB
XX are useful for treating Streptococcus pyogenes infections. The protein
XX is also useful for purifying alpha2M from a sample. The present sequence
XX is a DNA encoding partial GRAB protein from S. pyogenes strain KTL3.
XX
XX The protein has alpha2M binding region and is useful in vaccine
XX composition.
XX
SQ Sequence 504 BP; 188 A; 97 C; 108 G; 111 T; 0 other;

Alignment Scores:
Pred. No.: 4,89e-13 Length: 504
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AAD00564 (1-504)
QY 1 SerASPAlaLeuGluAlaLeuAlaAspGlnThrASPAlaLeuGlnSerGluGluAlaAla 20
DB 208 TCAGATGCGCTTGAAGCATTAGCGGATCAAGACAGACGTTTACAAATCAGAGAAAGCTGCG 267
QY 21 ValValIysAlaIAspAsnAlaAla 28
DB 268 GTTGTTAAGCGGATTAACGCTGCT 291

RESULT 3
AAD00560
ID AAD00560 standard; DNA: 654 BP.
XX
XX AAD00560;
XX
XX 29-AUG-2000 (first entry)
XX
XX Streptococcus pyogenes strain SF370 GRAB protein coding region.
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
XX alpha2-macroglobulin; group A Streptococcus; GAS; antidiabetic;
XX immune response; Streptococcus pyogenes infection; ds.
XX
XX Streptococcus pyogenes.
XX
XX Key Location/Qualifiers
XX
```

```

FH Key Location/Qualifiers
FT CDS 1..654
FT /tag- a
FT /product- "GRAB protein"
FT sig_peptide 1..99
FT /tag- b
FT mat_peptide 100..651
FT /tag- c
FT /product- "Mature GRAB protein"
XX
PN WO200026240-A2.
XX
PD 11-MAY-2000.
XX
PF 02-NOV-1999; 99MO-GB03631.
XX
PR 02-NOV-1998; 98GB-0023975.
XX
PA (ACT1-) ACTINOVA LTD.
XX
PI Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
DR P-PSDB; AAY71042.
XX
XX New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein
XX
XX Claim 13; Page 63; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.
XX
SQ Sequence 654 BP; 234 A; 118 C; 143 G; 159 T; 0 other;
XX
Alignment Scores:
Pred. No.: 6,79e-13 Length: 654
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
XX
US-09-847-539a-6_COPY_59_86 (1-28) x AAD00560 (1-654)
OY 1 SeraspalaLeuGluaLaLeuAlaAspGlnThraspalaLeuGlnSerGluAlaAla 20
Db 274 TCAGATGCGCTTAGAAGCATTAGCGGATCAACACACGCTTACATCAGAAAGAGCTGCG 333
OY 21 ValValysAlaAspAsnAlaAla 28
Db 334 GTTGTAAAGCGATTAAAGCGTCT 357
XX
RESULT 4
AAD00559
ID AAD00559 standard; DNA; 764 BP.
AC AAD00559;
XX
XX 29-AUG-2000 (first entry)
XX
DE Streptococcus pyogenes strain SF370 GRAB protein encoding DNA.
XX
KW GRAB protein: protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection; ds.
```

```

XX
XX Streptococcus pyogenes.
OS
XX
FH Key Location/Qualifiers
FT CDS 60..713
FT /tag- a
FT /product- "GRAB protein"
FT sig_peptide 60..158
FT /tag- b
FT mat_peptide 159..710
FT /tag- c
FT /product- "Mature GRAB protein"
FT primer_bind complement (101..124)
FT /tag- d
FT primer_bind complement (101..127)
FT /tag- e
FT primer_bind complement (160..184)
FT /tag- f
FT primer_bind complement (363..594)
FT /tag- g
FT primer_bind complement (563..594)
FT /tag- h
FT primer_bind complement (505..626)
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99MO-GB03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACT1-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
DR P-PSDB; AAY71042.
XX
XX New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein
XX
XX Example 1; Fig 2B; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.
XX
SQ Sequence 764 BP; 279 A; 131 C; 159 G; 195 T; 0 other;
XX
Alignment Scores:
Pred. No.: 8,27e-13 Length: 764
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
XX
US-09-847-539a-6_COPY_59_86 (1-28) x AAD00559 (1-764)
OY 1 SeraspalaLeuGluaLaLeuAlaAspGlnThraspalaLeuGlnSerGluAlaAla 20
Db 333 TCAGATGCGCTTAGAAGCATTAGCGGATCAACACACGCTTACATCAGAAAGAGCTGCG 392
```





```

OY      1 SerAspAlaLeuGluaIaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
DB      556 AGTACGCCCTTACAGATTGGCGGATCAACAGACGCTTACATCAGAGAAGAGCTGCG 615
OY      21 ValValIysAlaAspAsnAlaAla 28
DB      616 GTTGTAAACGGATACGCTGCT 639
RESULT 7
AAT67618
ID      AAT67618 standard; DNA; 576 BP.
XX
XX      AAT67618:
XX
DE      10-JUL-1997 (first entry)
XX
XX      H. pylori secreted or periplasmic protein ORF 35336707.aa.
XX
KM      Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KM      identification; binding compound; bacterium; life cycle; activator;
KM      bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KM      diagnosis; ds.
XX
XX      Helicobacter pylori.
XX
XX      Key      Location/Qualifiers
FH      CDS      1..576
FT      /*tag= a
FT      /transl_except= (pos: 541..543, aa: Glu)
FT      /transl_except= (pos: 547..549, aa: Asp)
FT      /note= "no stop codon given"
FT      559..561
FT      /*tag= b
FT      /note= "encodes Asn"
FT      574..576
FT      /*tag= c
FT      /note= "encodes Asn"
XX
XX      W09640893-A1.
XX
XX      19-DEC-1996.
XX
XX      06-JUN-1996; 96MO-US09122.
XX
XX      01-APR-1996; 96US-0630405.
XX      07-JUN-1995; 95US-0487032.
XX
XX      (ASTR ) ASTRA AB.
XX
XX      Berglindh OT, Smith D, Mellgaerd BL;
XX
XX      WPI, 1997-052306/05.
XX      P-PSDB; AAW20445.
XX
XX      Helicobacter pylori nucleic acid sequences and related
XX      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
XX      infection, and to detect Helicobacter
XX
XX      Claim 23; Page 268; 1481pp; English.
XX
XX      This sequence encodes a H. pylori secreted or periplasmic protein.
XX      The protein may be used in a vaccine to prevent or treat H. pylori
XX      infection or to identify H. pylori polypeptide binding compounds,
XX      useful as potential H. pylori life cycle activators or inhibitors.
XX      The genomic sequence of H. pylori (ATCC 55679) was determined from
XX      overlapping contigs generated by mechanically shearing the bacterial
XX      DNA. The sequences were analysed for ORF of at least 180 nucleotides,
XX      and the predicted coding regions defined by computer evaluation. To
XX      identify likely H. pylori antigens for vaccine development, the amino
XX      acid sequences predicted from various ORF were analysed for significant
XX      homology to other known or exported membrane proteins. Having identified
XX      and determined the sequences of interest, particular regions can be
XX      isolated from H. pylori by PCR amplification for recombinant polypeptide

```

```

CC      production, e.g. in E. coli hosts.
XX
SQ      Sequence 576 BP; 217 A; 97 C; 95 G; 161 T; 6 other:
XX
Alignment Scores:
Pred. No.:      1.02      Length:      576
Score:          57.00      Matches:      12
Percent Similarity: 69.57%      Conservative: 4
Best Local Similarity: 52.17%      Mismatches: 7
Query Match:     44.19%      Indels:      0
DB:             18      Gaps:         0
US-09-847-539a-6_copy_59_86 (1-28) x AAT67618 (1-576)
OY      4 LeuCluaIaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValIys 23
DB      145 ATACACGCCCTTACAGATTGGCGGATCAACAGACGCTTACATCAGAGAAGAGCTGCG 204
OY      24 AlaAspAsn 26
DB      205 TGGGATTAAC 213
RESULT 8
AAT67891
ID      AAT67891 standard; DNA; 666 BP.
XX
XX      AAT67891:
XX
DE      14-JUL-1997 (first entry)
XX
XX      H. pylori secreted or periplasmic protein ORF 02ce10216orfl.
XX
KM      Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KM      identification; binding compound; bacterium; life cycle; activator;
KM      bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KM      diagnosis; ds.
XX
XX      Helicobacter pylori.
XX
XX      Key      Location/Qualifiers
FH      CDS      1..666
FT      /*tag= a
FT      /note= "no stop codon given"
XX
XX      W09640893-A1.
XX
XX      19-DEC-1996.
XX
XX      06-JUN-1996; 96MO-US09122.
XX
XX      01-APR-1996; 96US-0630405.
XX      07-JUN-1995; 95US-0487032.
XX
XX      (ASTR ) ASTRA AB.
XX
XX      Berglindh OT, Smith D, Mellgaerd BL;
XX
XX      WPI, 1997-052306/05.
XX      P-PSDB; AAW20638.
XX
XX      Helicobacter pylori nucleic acid sequences and related
XX      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
XX      infection, and to detect Helicobacter
XX
XX      Claim 23; Page 770-771; 1481pp; English.
XX
XX      The present sequence encodes a H. pylori secreted or periplasmic protein.
XX      The protein may be used in a vaccine to prevent or treat H. pylori
XX      infection or to identify H. pylori polypeptide binding compounds,
XX      useful as potential H. pylori life cycle activators or inhibitors.
XX      The genomic sequence of H. pylori (ATCC 55679) was determined from
XX      overlapping contigs generated by mechanically shearing the bacterial
XX      DNA. The sequences were analysed for ORF of at least 180 nucleotides,

```

CC and the predicted coding regions defined by computer evaluation. No  
CC identity likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.

XX SQ Sequence 666 BP; 250 A; 115 C; 119 G; 182 T; 0 other;

Alignment Scores:  
Pred. No.: 1,23 Length: 666  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 18 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AAT67891 (1-666)

OY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValys 23  
DB 229 ATACAAAGCCCTACAGAGCAATTCAGCTTACATTCACAGAAAAGTCCTTACGAAA 288

OY 24 AlaAspAsn 26  
DB 289 TGGGATTAAC 297

RESULT 9  
AAK30461  
ID AAK30461 standard; DNA: 1239 BP.  
XX AAK30461;  
XX AC AAK30461;  
XX DT 08-JUN-1999 (first entry)  
XX DE H. pylori secreted protein ORF 09cel0413\_35336707\_f2.9.  
XX KW Vaccine: probe: diagnostic; ORF; cell envelope protein;  
XX KM secreted protein; cellular protein; ds.  
XX OS Helicobacter pylori.  
XX PN W09818323-A1.  
XX PD 07-MAY-1998.  
XX PF 28-OCT-1997; 97WO-US19575.  
XX PR 14-JUL-1997; 97US-0891928.  
XX PR 28-OCT-1996; 96US-0739150.  
XX PR 06-DEC-1996; 96US-0759739.  
XX PA (ASTR ) ASTRA AB.  
XX PI Alm RA, Smith D;  
XX DR WPI: 1998-271811/24.  
XX DR P-PSDB: AAY10994.

XX Helicobacter pylori nucleic acids and proteins - used to develop  
PT products for the detection, prevention and treatment of H. pylori  
PT infections  
XX Claims 3, 4; Page 126-127; 279pp; English.

XX Recombinant or substantially pure preparations of H. pylori polypeptides  
CC are disclosed, together with the nucleic acids encoding them. In all,  
CC 73 ORFs are shown. The proteins are variously cell envelope proteins,  
CC secreted proteins or other cellular proteins. Vaccines containing the  
CC nucleic acids or proteins are claimed, as are probes containing at least  
CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful  
CC for treating or reducing the risk of H. pylori infections, and the

CC probes can be used diagnostically for detecting the presence of  
CC Helicobacter in a sample. The products are also of use in screening  
CC for compounds having the ability to interfere with the H. pylori life  
CC cycle or to inhibit H. pylori infection.

XX SQ Sequence 1239 BP; 472 A; 234 C; 206 G; 327 T; 0 other;

Alignment Scores:  
Pred. No.: 2,69 Length: 1239  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 19 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AAK30461 (1-1239)

OY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValys 23  
DB 145 ATACAAAGCCCTACAGAGCAATTCAGCTTACATTCACAGAAAAGTCCTTACGAAA 204

OY 24 AlaAspAsn 26  
DB 205 TGGGATTAAC 213

RESULT 10  
ABL13527/C  
ID ABL13527 standard; CDNA: 3349 BP.  
XX ABL13527;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 35063.  
XX KW Drosophila: developmental biology; cell signalling; insecticide;  
XX KM pharmaceutical; gene; ss.  
XX OS Drosophila melanogaster.  
XX PN W0200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US09231.  
XX PR 23-MAR-2000; 2000US-191637P.  
XX PR 11-JUL-2000; 2000US-0614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EM;  
XX DR WPI: 2001-656860/75.  
XX DR P-PSDB: ABB69424.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX Claim 1; SEQ ID NO 35063; 21pp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL6175) and the encoded proteins  
CC (AB57737-AB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 3349 BP; 808 A; 953 C; 851 G; 737 T; 0 other;  
Alignment Scores:  
Pred. No.: 30.6 Length: 3349  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 23 Gaps: 0  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x ABL13527 (1-3349)  
OY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIaValIysAla 24  
DB 1122 GAGGAGGAAGCTGACGAGAGACGCCGCCGAGCAGATGCCGCGCTGCCGCC 1063  
OY 25 AspasnAlaala 28  
DB 1062 GATGCTGCGCGC 1051  
RESULT 11  
ABL13526/c  
ID ABL13526 standard; cDNA; 6398 BP.  
XX ABL13526;  
AC  
XX 26-MAR-2002 (first entry)  
DT  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 35060.  
DE  
XX Drosophila: developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX W0200171042-A2.  
PN  
XX  
XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PE  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX 11-JUL-2000; 2000US-0614150.  
PR  
XX (PEKE ) PE CORP NY.  
PA  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX  
XX WPI: 2001-656860/75.  
DR  
XX P-PSDB; ABB69423.  
DR  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX  
XX Claim 1; SEQ ID NO 35060; 21pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL10511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pat\_sequences.  
CC  
XX  
XX Sequence 6398 BP; 1688 A; 1632 C; 1352 G; 1726 T; 0 other;  
SQ  
Alignment Scores:

Pred. No.: 69.3 Length: 6398  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 23 Gaps: 0  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x ABL13526 (1-6398)  
OY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIaValIysAla 24  
DB 2298 GAGGAGGAAGCTGACGAGAGACGCCGCCGAGCAGATGCCGCGCTGCCGCC 2239  
OY 25 AspasnAlaala 28  
DB 2238 GATGCTGCGCGC 2227  
RESULT 12  
AAC54228/c  
ID AAC54228 standard; DNA; 1002 BP.  
XX AAC54228;  
AC  
XX 18-OCT-2000 (first entry)  
DT  
XX  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 77144.  
DE  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX EP1033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PE  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR  
XX 05-MAR-1999; 99US-0123180.  
PR  
XX 09-MAR-1999; 99US-0123548.  
PR  
XX 23-MAR-1999; 99US-0125768.  
PR  
XX 25-MAR-1999; 99US-0126264.  
PR  
XX 29-MAR-1999; 99US-0126785.  
PR  
XX 01-APR-1999; 99US-0127462.  
PR  
XX 06-APR-1999; 99US-0128234.  
PR  
XX 08-APR-1999; 99US-0128714.  
PR  
XX 16-APR-1999; 99US-0129845.  
PR  
XX 19-APR-1999; 99US-0130077.  
PR  
XX 21-APR-1999; 99US-0130449.  
PR  
XX 23-APR-1999; 99US-0130510.  
PR  
XX 28-APR-1999; 99US-0130891.  
PR  
XX 30-APR-1999; 99US-0131449.  
PR  
XX 30-APR-1999; 99US-0132048.  
PR  
XX 30-APR-1999; 99US-0132407.  
PR  
XX 04-MAY-1999; 99US-0132484.  
PR  
XX 05-MAY-1999; 99US-0132485.  
PR  
XX 06-MAY-1999; 99US-0132486.  
PR  
XX 07-MAY-1999; 99US-0132487.  
PR  
XX 11-MAY-1999; 99US-0132863.  
PR  
XX 14-MAY-1999; 99US-0134218.  
PR  
XX 14-MAY-1999; 99US-0134219.  
PR  
XX 14-MAY-1999; 99US-0134221.  
PR  
XX 14-MAY-1999; 99US-0134370.  
PR  
XX 18-MAY-1999; 99US-0134766.  
PR  
XX 19-MAY-1999; 99US-0134941.  
PR  
XX 20-MAY-1999; 99US-0135124.  
PR  
XX 21-MAY-1999; 99US-0135353.  
PR  
XX 24-MAY-1999; 99US-0135629.  
PR  
XX 25-MAY-1999; 99US-0136021.  
PR  
XX 27-MAY-1999; 99US-0136392.  
PR



Alignment Scores:  
Pred. No.: 14.6  
Score: 52.00  
Percent Similarity: 72.73%  
Best Local Similarity: 50.00%  
Query Match: 40.31%  
DB: 21  
Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AAC54228 (1-1002)

Qy 4 leuGluaAlaLeuAlaaspGlnThrAspAlaLeuGlnSerGluAlaAlaValIlys 23  
Db 442 CTTGCCGCGACTCTCCGACCAAGGACATGCGCTCAATTCAAAGCTGCAATGCGCAAG 383  
Qy 24 Alaasp 25  
Db 382 GCGGAT 377

RESULT 13  
AAC33266/C  
ID AAC33266 standard; DNA; 1230 BP.  
XX  
AC AAC33266;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 2396.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KM protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 18-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137502.  
PR 08-JUN-1999; 99US-0137724.  
PR 10-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148341.  
PR 12-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158294.  
PR 13-OCT-1999; 99US-0159229.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.

PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Alignment Scores:  
Pred. No.: 18.9  
Score: 52.00  
Percent Similarity: 72.73%  
Best Local Similarity: 50.00%  
Query Match: 40.31%  
DB: 21  
Gaps: 0  
  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x AAC33266 (1-1230)  
Oy 4 LeuGLuAlaLeuAlaAspLInThrasPaLaLeuGInserGluGluAlaAlaValLys 23  
||| |||||:::||||: ||| ||| :::::|||||||::: |||  
Db 518 CTTCCCGCACACTCTCCGACAAAGGACATGCCCTCAATTCMAAAGCTGCATCGCAAG 459  
  
Oy 24 AlaAsp 25  
|||||  
Db 458 GCGGAT 453  
  
RESULT 14  
ABL25211/c  
ID ABL25211 standard; DNA; 16135 BP.  
XX  
AC ABL25211;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 27106.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ds.  
XX  
CS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PMD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 27106; 21bp + sequence listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB16351), expressed DNA  
CC sequences (AB161840-AB16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 16135 BP; 4290 A; 3936 C; 4034 G; 3875 T; 0 other;  
  
Alignment Scores: 488 Length: 16135  
Pred. No.: 16135

```

Score:          52.00          Matches:          11
Percent Similarity: 65.22%      Conservative: 4
Best Local Similarity: 47.83%    Mismatches: 8
Query Match:      40.31%        Indels: 0
DB:               23           Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL25210 (1-16135)

OY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLys 23
    |||||
Db 5971 CTGGAGAGCTCTGTTCAAGCAGTGTCTCTCAAGCCGACGTTGCTCTTGGCA 5912
    |||||
OY 24 AlaAspAsn 26
    |||||
Db 5911 GCCGACAAAT 5903

RESULT 15
ABL25210
ID ABL25210 standard; DNA; 18603 BP.
XX
AC ABL25210;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 27103.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
XX
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 27103; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (AB01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS7072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 18603 BP; 4763 A; 4409 C; 4291 G; 5140 T; 0 other;

Alignment Scores:
Pred. No.: 584          Length: 18603
Score: 52.00          Matches: 11
Percent Similarity: 65.22%      Conservative: 4
Best Local Similarity: 47.83%    Mismatches: 8
Query Match: 40.31%        Indels: 0
DB: 23           Gaps: 0

```

```

US-09-847-539A-6_COPY_59_86 (1-28) x ABL25210 (1-18603)

OY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLys 23
    |||||
Db 11563 CTGGAGAGCTCTGTTCAAGCAGTGTCTCTCAAGCCGACGTTGCTCTTGGCA 11622
    |||||
OY 24 AlaAspAsn 26
    |||||
Db 11623 GCCGACAAAT 11631

RESULT 16
ABL01876/c
ID ABL01876 standard; cDNA; 25461 BP.
XX
AC ABL01876;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 110.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
XX
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR P-PSDB; ABBS7773.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 110; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (AB01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS7072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 25461 BP; 7542 A; 5394 C; 5424 G; 7101 T; 0 other;

Alignment Scores:
Pred. No.: 1,06e+03          Length: 25461
Score: 51.50          Matches: 13
Percent Similarity: 65.38%      Conservative: 4
Best Local Similarity: 50.00%    Mismatches: 8
Query Match: 39.92%        Indels: 1
DB: 23           Gaps: 1

US-09-847-539A-6_COPY_59_86 (1-28) x ABL01876 (1-25461)

OY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21

```

```

Db 12848 GACCGCTTTAAAGCGCTTTCTTCACGACGTAT--TTGAGCCATCCGATGATGATAT 12792
Oy 22 VallysAlaAspAsnAla 27
Db 12791 TTTAAGCAGACACATGCC 12774

RESULT 17
ABLI2683/c
ID ABLI2683 standard; cDNA; 1960 BP.
XX
XX ABLI2683:
XX
XX AC
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 32531.
XX KW Drosophila: developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX DR P-PSDB; ABB68580.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX interactions -
XX
XX PS Claim 1; SEQ ID NO 32531; 21pp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX sequences (ABLI01840-ABLI6175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX CC The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcc_sequences.
XX
XX SQ Sequence 1960 BP; 505 A; 526 C; 485 G; 444 T; 0 other;

Alignment Scores:
Pred. No.: 50.3 Length: 1960
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABLI2683 (1-1960)
Oy 2 AspaAlaLeuGluAlaLeuAlaAspGlnThrAspaAlaLeuGlnSerGluGluAlaAlaVal 21
Db 228 GATCGCGGCAATCATATATGAAAGCACACAGACGAGAGAGACGAGACAAACAGCGCGC 169
Oy 22 VallysAlaAspAsnAla 27

```

```

Db 168 GTACGAGAGATCGTGC 151
Oy 22 VallysAlaAspAsnAla 27
Db 168 GTACGAGAGATCGTGC 151

RESULT 18
ABLI27360
ID ABLI27360 standard; DNA; 2618 BP.
XX
XX ABLI27360:
XX
XX AC
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 33553.
XX KW Drosophila: developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX DR
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX PS Claim 1; SEQ ID NO 33553; 21pp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX sequences (ABLI01840-ABLI6175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX CC The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcc_sequences.
XX
XX SQ Sequence 2618 BP; 669 A; 603 C; 652 G; 694 T; 0 other;

Alignment Scores:
Pred. No.: 72.5 Length: 2618
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABLI27360 (1-2618)
Oy 2 AspaAlaLeuGluAlaLeuAlaAspGlnThrAspaAlaLeuGlnSerGluGluAlaAlaVal 21
Db 467 GATCGCGGCAATCATATATGAAAGCACACAGACGAGAGAGACGAGACAAACAGCGCGC 526
Oy 22 VallysAlaAspAsnAla 27
Db 527 GTACGAGAGATCGTGC 544
Oy 22 VallysAlaAspAsnAla 27

```





RESULT 21  
AAS94455  
ID AAS94455 standard; cDNA; 3711 BP.

XX AC AAS94455;  
XX DT 13-FEB-2002 (first entry)  
XX DE DNA encoding novel human diagnostic protein #30259.  
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX OS Homo sapiens.  
XX PN M02001.75067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001MO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI: 2001-639362/73.  
XX DR P-PsDB: ABG30268.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -  
XX PS Claim 1; SEQ ID No 30259; 103pp: English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. M564197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from Wipo  
CC at [http://wipo.int/pub/published\\_pat\\_sequences](http://wipo.int/pub/published_pat_sequences).  
XX SO Sequence 3711 BP; 848 A; 945 C; 1056 G; 862 T; 0 other;

Alignment Scores:  
Pred. No.: 167  
Percent Similarity: 50.00 Length: 3711  
Best Local Similarity: 66.67% Matches: 11  
Query Match: 45.83% Mismatches: 8  
DB: 38.76% Indels: 0  
Gaps: 23

SO-09-847-539A-6\_COPY\_59\_86 (1-28) x AAS94455 (1-3711)



APPLICATION NUMBER: US/08/591.079  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Livnat, Shmuel  
REGISTRATION NUMBER: 33,949  
REFERENCE/DOCKET NUMBER: 15661-20017.00  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0764  
TELEX: 90-4030 MRSNFOERSWSH  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5393 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: Salmonella typhi  
STRAIN: Ty2  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 543..2324  
OTHER INFORMATION: /gene= "sfpB"  
US-08-591-079-9

Alignment Scores:  
Pred. No.: 13.3 Length: 5393  
Score: 66.00 Matches: 23  
Percent Similarity: 40.70% Conservative: 12  
Best Local Similarity: 26.74% Mismatches: 51  
Query Match: 8.43% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-591-079-9 (1-5393)

Qy 58 SerSerAspAlaLeuGluAlaLeuAlaLeuGlnThrAspAlaLeuGlnSerGluAla 77  
Db 957 CGGCGGCTCTATGATCCAGCATCAAAAGAGCGGATGATCCGCAAGAGTTTATGAC 1016  
Qy 78 AlavalValLysAlaAspAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThr 97  
Db 1017 GCTGGCGCAAAACTGACGCGGCAAAATAATTCGATCGTGACCCAGCTGAC 1076  
Qy 98 AspAlaLeuGlnSerGluGluAlaGluValGlnSerAspAlaLeuAlaSerAspAla 117  
Db 1077 CCGGCTATGCACAGCTCAAGCCCGCTAGAACACAGCGGGAAGAGAGCGAGAGGCG 1136  
Qy 118 TtpGluLysAlaLeuThrProfileAlaLeuAspValLysThrLysAspThrLysPro 137  
Db 1137 AAGAGCGCTTAGATAGCCAGGATGCGGCGTTAAGCAGGACAGACGCCAAGGG 1196  
Qy 138 ValValLysLysGluGlu 143  
Db 1197 AAAGCGGAGAAAGCGGAT 1214

RESULT 16

US-08-728-323A-1  
Sequence 1, Application US/08728323A  
Patent No. 5948676

GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
NUMBER OF INVENTIONS: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728.323A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3489 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3489  
US-08-728-323A-1

Alignment Scores:  
Pred. No.: 10.5 Length: 3489  
Score: 65.00 Matches: 18  
Percent Similarity: 45.69% Conservative: 35  
Best Local Similarity: 15.52% Mismatches: 63  
Query Match: 8.37% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-728-323A-1 (1-3489)  
Qy 30 LeuArgAsnGluGluArgAlaLeuLeuLeuLysGlnAlaLeuGluAspLysGlu 49  
Db 2326 TTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGGAGCAG 2385  
Qy 50 AlaThrThrAlaLeuAlaLeuAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69  
Db 2386 GAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAG 2445  
Qy 70 AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAlaAlaSerAspAla 89  
Db 2446 CAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGCAGGAG 2505  
Qy 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln 109  
Db 2506 GAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGCAGGAGGAG 2565  
Qy 110 SerAspAlaLeuAlaSerAspAlaThrGluLysAlaAlaThrProfileAlaLeuAspVal 129  
Db 2566 CAGGAGCAGGAGCAGGAGGAGTTAGAGGAGGAGTTAGAGGAGCAGGAGCAGGAGGAGGAG 2625  
Qy 130 LysLysThrLysAspThrLysProValValLysLysLysGluGluArgGln 145  
Db 2626 CAGGAGGAGCAGGAGTTAGAGGAGGAGTTAGAGGAGGAGTTAGAGGAGCAGGAGCAGGAG 2673

RESULT 17

US-08-770-379-20/c  
Sequence 20, Application US/08770379  
Patent No. 5849564  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan

```
;
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1419 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1416
; ORGANISM: Salmonella typhimurium
; STRAIN: sibb
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 575..2356
; OTHER INFORMATION: /product= "SipB"
; OTHER INFORMATION: /gene= "sipB"
;
; US-08-591-079-7
;
; Alignment Scores:
; Pred. No.: 1.57 Length: 1419
; Score: 67.00 Matches: 18
; Percent Similarity: 42.22% Conservative: 20
; Best Local Similarity: 20.00% Mismatches: 52
; Query Match: 8.62% Indels: 0
; DB: 4 Gaps: 0
;
; US-09-847-539A-6 (1-159) x US-09-115-746-9 (1-1419)
;
; Qy 33 GluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThr 52
; |||:||||| ||| ||| |||:|||||
; Db 763 GAGAAGCAGAGCGGCGTGAAGCCAGCAGAGGTGCGGAAGCGGAGCAGCAGAGCGCT 822
;
; Qy 53 AlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 72
; |||: ||| |||: |||
; Db 823 GAAGCCAGCAAGGTGCGGAAGCGGAGCAGAGCGGCGTGAAGCCAGCAGAGGTGCC 882
;
; Qy 73 GlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAla 92
; |||: ||| ||| ||| |||: |||
; Db 883 GAAGCGGAGAGCAAGCGCAGCTGAGCCACGAGGTGCGGAGCAGCAGCAGCAGAG 942
;
; Qy 93 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsn 112
; |||: ||| |||: ||| |||: |||
; Db 943 GCAGCTGAAGCCAGCAGAGGTGCGGAAGCGGAGCAGAGCAGCAGCTGAAGCCAGCAG 1002
;
; Qy 113 AlaAlaSerAspAlaTrpGluLysAlaAla 122
; ||| |||: ||| ||| |||
; Db 1003 GTTCCGAAGCGGAGCAGCAGAGCAGCT 1032
;
; RESULT 14
; US-08-591-079-7
; Sequence 7, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: Zychlinsky, Arturo
; APPLICANT: Chen, Yajing
; TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591.079
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Livnat, Shmuel
; REGISTRATION NUMBER: 33,949
;
; REFERENCE/DOCKET NUMBER: 15661-20017.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0764
; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3622 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Salmonella typhimurium
; STRAIN: sibb
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 575..2356
; OTHER INFORMATION: /product= "SipB"
; OTHER INFORMATION: /gene= "sipB"
;
; US-08-591-079-7
;
; Alignment Scores:
; Pred. No.: 7.8 Length: 3622
; Score: 66.00 Matches: 23
; Percent Similarity: 40.70% Conservative: 12
; Best Local Similarity: 26.74% Mismatches: 51
; Query Match: 8.49% Indels: 0
; DB: 2 Gaps: 0
;
; US-09-847-539A-6 (1-159) x US-08-591-079-7 (1-3622)
;
; Qy 58 SerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 77
; |||: ||| ||| ||| |||: |||
; Db 989 GCGAGCGGATCTCTATGAAGCCAGTATCAAAAGAGCAGGATACCGCCAGAGTGTATTATGAC 1048
;
; Qy 78 AlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThr 97
; ||| ||| ||| ||| |||: |||
; Db 1049 GCTCGCAGCAAAACTGACGCGCGCAAAATAAATTGCAATCGCTGGACCCGCGCTGAC 1108
;
; Qy 98 AspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAla 117
; ||| ||| ||| ||| |||: |||
; Db 1109 CCGGCTATGCACAGCTGAAGCCGCGGTAGAACGCCGCGGAGGAGGAGGAGGAGGAGG 1168
;
; Qy 118 TrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysPro 137
; ||| ||| ||| ||| |||: |||
; Db 1169 AAAGAGCGCTTAGATAGCGCCAGCGATCGAGCGGTTAAAGCAGGCGACAGCCCAAGCG 1228
;
; Qy 138 ValValLysLysGluGlu 143
; ||| ||| ||| |||
; Db 1229 AAAGCGGAGAAAGCGGAT 1246
;
; RESULT 15
; US-08-591-079-9
; Sequence 9, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: Zychlinsky, Arturo
; APPLICANT: Chen, Yajing
; TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591.079
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Livnat, Shmuel
; REGISTRATION NUMBER: 33,949
;
```



;; TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
;; TITLE OF INVENTION: Polypeptides and their Uses  
;; NUMBER OF SEQUENCES: 534  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Smithline Beecham Corporation  
;; STREET: 709 Swedeland Road  
;; CITY: King of Prussia  
;; STATE: PA  
;; COUNTRY: USA  
;; ZIP: 19406-0939  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FASTSEQ for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/936.165A  
;; FILING DATE: 24-SEP-1997  
;; CLASSIFICATION: 536

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 60/027.032  
;; FILING DATE: 24-SEP-1996

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Gimmi, Edward R

;; REGISTRATION NUMBER: 38,891

;; REFERENCE/DOCKET NUMBER: P50549

;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 610-270-4478  
;; TELEFAX: 610-270-5090

;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 93:

;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 605 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: Genomic DNA

US-08-936-165A-93

Alignment Scores:

Pred. No.: 0.127 Length: 605

Score: 71.00 Matches: 20

Percent Similarity: 41.49% Conservativeness: 19

Best Local Similarity: 21.28% Mismatches: 55

Query Match: 9.14% Indels: 0

DB: 4 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-936-165A-93 (1-605)

QY 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 81

DB 588 CTCGAGGCATGCAAGCTTGAGTCTTCTNTAATTCAGACACACACACAGCAGT 529

QY 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101

DB 528 AATCGAATGCGAAATGAAGTTCATCAAGCTGTGACAACTCAAAATCAAGCAATTGAT 469

QY 102 SerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAla 121

DB 468 ATACAACCTGGCGTACAACTGAGAGAAATGAGCAAAAGATTAGTTTANAAGCT 409

QY 122 AlaThrProileAlaLeuAlaValLysThrLysProValVallysLys 141

DB 408 AAGAAAAAGCGTNTCAAGATATCTTAATGTCACAAACAACTAATGATGTTACGCAATT 349

QY 142 GluGluArgGlnAsnValAsnThrLeuProThrGlyGlu 155

DB 348 AATGATCAAGCAGTTCCTGCTATATCAAGGTATTACTGGAGAT 307

RESULT 10

US-08-973-462-2

; Sequence 2, Application US/08973462B

; Patent No. 6191270

;; GENERAL INFORMATION:  
;; APPLICANT: DAUBERSIES, PIERRE  
;; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
;; FILE REFERENCE: 0660-0125-0 PCT  
;; CURRENT APPLICATION NUMBER: US/08/973.462B  
;; EARLIER FILING DATE: 1998-02-06  
;; EARLIER FILING DATE: 1996-06-12  
;; EARLIER FILING DATE: 1995-06-13  
;; NUMBER OF SEQ ID NOS: 29  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 2  
;; LENGTH: 5361  
;; TYPE: DNA  
;; ORGANISM: P. falciparum  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1)..(5361)  
US-08-973-462-2

Alignment Scores:

Pred. No.: 6.66 Length: 5361

Score: 68.00 Matches: 17

Percent Similarity: 42.06% Conservativeness: 28

Best Local Similarity: 15.89% Mismatches: 62

Query Match: 8.75% Indels: 0

DB: 4 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-973-462-2 (1-5361)

QY 26 GluLysLeuAlaLeuArgAspGluGluArgAlaIleAspGluLysLysGlnAlaIle 45

DB 682 GAAGAACTGTAGAGAAATGACCAAGAAAGTGTAGAAGAAATGTAGAAGAAATGTA 741

QY 46 GluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu 65

DB 742 GAAGAAATGACGACGAAAGTGTAGCTCAAGTGTGAAGAAAGTATAGCTTCAAGTGT 801

QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 85

DB 802 GATGAAGTATAGATTCAAGTATTGAGAAATGATGCTCAACTGTTGAGAAATGCTGA 861

QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105

DB 862 GCTCAAGTGTGTAGAAGTGTGCTCCAACTGTTGAAGAAAGTGTAGAAGAAATGTT 921

QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProile 125

DB 922 GAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGA 981

QY 126 AlaLeuAspValLysLysThr 132

DB 982 GCTGAAATGTTGAAGAAAGT 1002

RESULT 11

US-08-973-462-1

; Sequence 1, Application US/08973462B

; Patent No. 6191270

;; GENERAL INFORMATION:  
;; APPLICANT: DAUBERSIES, PIERRE  
;; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
;; FILE REFERENCE: 0660-0125-0 PCT  
;; CURRENT APPLICATION NUMBER: US/08/973.462B  
;; EARLIER FILING DATE: 1998-02-06  
;; EARLIER FILING DATE: 1996-06-12  
;; EARLIER FILING DATE: 1995-06-13  
;; NUMBER OF SEQ ID NOS: 29  
;; SOFTWARE: PatentIn Ver. 2.0







```
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORGANISM: HUMAN
US-08-353-700-2

Alignment Scores:
Pred. No.: 1.01 Length: 10136
Score: 76.00 Matches: 21
Percent Similarity: 49.00% Conservatve: 28
Best Local Similarity: 21.00% Mismatches: 51
Query Match: 9.78% Indels: 0
DB: 1 Gaps: 0

US-09-847-539a-6 (1-159) x US-08-353-700-2 (1-10136)
QY 46 GluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu 65
Db 7728 GAAAGAGCAAGAGCAAGAGTACAGATGAAGAAATAATCAAGCACTGCCATGGAGATGCTT 7787
QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAla 85
Db 7788 CAACACAAATTAAGAGCTCAATGAGAGAGTGGCAGCCCTGCATATACCAAGAGCC 7847
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
Db 7848 TGTAGGCGCAAGAGCAAGATCTTAGTAGTCAAGTAGAGTGTCTTCAACTGAGAGGCT 7907
QY 106 GluValValGlnSerAspAlaAlaSerAspAlaTlpGluLysAlaAlaThrProIle 125
Db 7908 CAGTTGCTACAGGCTTGTATGAGGCAAAATAATATATATTTTGCATCTTCAGTG 7967
QY 126 AlaLeuAspValLysThrLysAspThrLysProValValLysLysGluGluArgGln 145
Db 7968 AAAGCCCTATTCAAGACTAGAGTGCACAGCAAGAACTGGAGAGAGGATGAAGAA 8027

RESULT 4
PCT-US95-16216-2
; Sequence 2, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-2

Alignment Scores:
Pred. No.: 1.01 Length: 10136
Score: 76.00 Matches: 21
Percent Similarity: 49.00% Conservatve: 28
Best Local Similarity: 21.00% Mismatches: 51
Query Match: 9.78% Indels: 0
DB: 5 Gaps: 0

US-09-847-539a-6 (1-159) x PCT-US95-16216-2 (1-10136)
QY 46 GluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu 65
Db 7728 GAAAGAGCAAGAGCAAGAGTACAGATGAAGAAATAATCAAGCACTGCCATGGAGATGCTT 7787
QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAla 85
Db 7788 CAACACAAATTAAGAGCTCAATGAGAGAGTGGCAGCCCTGCATATACCAAGAGCC 7847
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
Db 7848 TGTAGGCGCAAGAGCAAGATCTTAGTAGTCAAGTAGAGTGTCTTCAACTGAGAGGCT 7907
QY 106 GluValValGlnSerAspAlaAlaSerAspAlaTlpGluLysAlaAlaThrProIle 125
Db 7908 CAGTTGCTACAGGCTTGTATGAGGCAAAATAATATATATTTTGCATCTTCAGTG 7967
QY 126 AlaLeuAspValLysThrLysAspThrLysProValValLysLysGluGluArgGln 145
Db 7968 AAAGCCCTATTCAAGACTAGAGTGCACAGCAAGAACTGGAGAGAGGATGAAGAA 8027

RESULT 5
US-08-145-705A-1/G
; Sequence 1, Application US/08145705A
; Patent No. 5489513
; GENERAL INFORMATION:
; APPLICANT: Springer, Wolfgang; Piempel, Manfred;
; APPLICANT: L.berding, Antonius
; TITLE OF INVENTION: SPECIFIC GENE PROBES AND
; TITLE OF INVENTION: PROCESSES FOR THE DIAGNOSTIC
; TITLE OF INVENTION: INVESTIGATION OF CANDIDA
; TITLE OF INVENTION: ALBICANS
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPRUNG HORN KRAMER & WOODS
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage
; COMPUTER: NEC PowerMate 1 Plus
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/145,705A
; FILING DATE: October 28, 1993
```

```
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 61743/102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1555 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 288..1526
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 288..1526
;
US-08-669-408B-9
Alignment Scores:
Pred. No.: 0.000481 Length: 1555
Score: 91.00 Matches: 19
Percent Similarity: 62.50% Conservative: 12
Best Local Similarity: 59.38% Mismatches: 12
Query Match: 11.71% Indels: 0
DB: 3 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-669-408B-9 (1-1555)
Qy 128 AspValLysLysThrLysAspThrLysProValValLysLysGluGluAArgGlnAsnVal 147
Db 1356 GATCTAGGAAGAGATGCTAAAAACAGAGCTAAGAAAGAAAGAGCTAAGAAAGCT 1415
Qy 148 AsnThrLeuProThrGlyGluGluSerAsnPro 159
Db 1416 GCAACTCTCTCACTGCTGGAAGGAAGCAACCCA 1451

RESULT 2
US-08-669-408B-1
; Sequence 1, Application US/08669408B
; Patent No. 6100055
; GENERAL INFORMATION:
; APPLICANT: GUSS, Bengt
; APPLICANT: JONSSON, Hans
; APPLICANT: LINDBERG, Martin
; APPLICANT: MUELLER, Hans-Peter
; APPLICANT: RANTAMAKI, Liisa K.
; TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING
; TITLE OF INVENTION: PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,408B
; FILING DATE: 03-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/SE94/00826
; FILING DATE: 06-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9302855-3
```

```
; FILING DATE: 06-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 61743/102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2526 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 506..2497
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 506..2497
;
US-08-669-408B-1
Alignment Scores:
Pred. No.: 0.00719 Length: 2526
Score: 85.00 Matches: 18
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 12
Query Match: 10.94% Indels: 0
DB: 3 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-669-408B-1 (1-2526)
Qy 128 AspValLysLysThrLysAspThrLysProValValLysLysGluGluAArgGlnAsnVal 147
Db 2327 GACGCTAAAAACAGAGCTAAGAAAGAAAGAGCTAAGAAAGAGCTAAGAAAGCT 2386
Qy 148 AsnThrLeuProThrGlyGluGluSerAsnPro 159
Db 2387 GCAACTCTCTCACTGCTGGAAGGAAGCAACCCA 2422

RESULT 3
US-08-353-700-2
; Sequence 2, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RAYNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:17:36 ; Search time 51.8663 Seconds  
(without alignments)  
753.008 Million cell updates/sec

Title: US-09-847-539A-6  
Perfect score: 777  
Sequence: 1 VDSIEQPIIPNGITLNL.....KKEERQNVNLTPTGERSNP 159

Scoring table:

BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 393533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09847539/runat\_10102002\_092549\_3456/app\_query.fasta\_1.526  
-DB=Issued\_Patents\_NA -QWTF=fastap -SUFFIX=rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539-ecgn\_1\_13 -runat\_10102002\_092549\_3456 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGEOUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	11.7	1555	3	US-08-669-408B-9
2	85	10.9	2526	3	US-08-669-408B-1
3	76	9.8	10136	1	US-08-353-700-2
4	76	9.8	10136	5	PCT-US95-16216-2
5	75	9.7	2233	1	US-08-145-705A-1
6	75	9.7	4868	1	US-08-139-937-12
7	75	9.7	4868	5	PCT-US93-11310-12
8	75	9.7	8789	1	US-08-328-254-5
9	71	9.1	605	4	US-08-936-165A-93
10	68	8.8	5361	4	US-08-973-462-2
11	68	8.8	6152	4	US-08-973-462-1
12	67	8.6	1419	2	US-08-216-894-9
					Sequence 9, Appl
					Sequence 1, Appl
					Sequence 2, Appl
					Sequence 12, Appl
					Sequence 5, Appl
					Sequence 93, Appl
					Sequence 2, Appl
					Sequence 1, Appl
					Sequence 9, Appl

13	67	8.6	1419	4	US-09-115-746-9	Sequence 9, Appl
14	66	8.5	2622	2	US-08-591-079-7	Sequence 7, Appl
15	66	8.5	5393	2	US-08-591-079-9	Sequence 9, Appl
16	65	8.4	3489	2	US-08-728-323A-1	Sequence 1, Appl
17	65	8.4	32207	2	US-08-770-379-20	Sequence 20, Appl
18	65	8.4	32207	4	US-08-757-669A-20	Sequence 20, Appl
19	65	8.4	32207	4	US-09-230-371A-20	Sequence 20, Appl
20	64	8.2	1695	2	US-08-216-894-1	Sequence 1, Appl
21	64	8.2	1695	4	US-09-115-746-1	Sequence 1, Appl
22	64	8.2	1932	2	US-08-216-894-7	Sequence 7, Appl
23	64	8.2	1932	4	US-09-115-746-7	Sequence 7, Appl
24	64	8.2	4316	1	US-08-317-450B-14	Sequence 14, Appl
25	64	8.2	4316	3	US-08-800-593-14	Sequence 14, Appl
26	64	8.2	5200	1	US-08-317-450B-12	Sequence 12, Appl
27	64	8.2	5200	3	US-08-800-593-12	Sequence 12, Appl
28	63	8.1	867	2	US-08-961-858-2	Sequence 2, Appl
29	63	8.1	867	3	US-09-961-858-4	Sequence 4, Appl
30	63	8.1	867	3	US-09-089-593-2	Sequence 2, Appl
31	63	8.1	867	3	US-09-089-593-3	Sequence 3, Appl
32	63	8.1	867	3	US-08-950-925-3	Sequence 3, Appl
33	62	8.0	397	3	US-09-253-691-3	Sequence 3, Appl
34	62	8.0	1890	4	US-09-346-408-7	Sequence 7, Appl
35	62	8.0	3252	2	US-08-770-301A-4	Sequence 4, Appl
36	62	8.0	4248	3	US-08-175-581-4	Sequence 4, Appl
37	62	8.0	4378	2	US-08-678-614-1	Sequence 1, Appl
38	62	8.0	4378	2	US-09-080-897-3	Sequence 3, Appl
39	62	8.0	4378	4	US-09-323-735-3	Sequence 3, Appl
40	62	8.0	4399	3	US-08-899-595-2	Sequence 2, Appl
41	62	8.0	6744	1	US-08-119-125A-2	Sequence 2, Appl
42	62	8.0	8878	1	US-08-759-444-2	Sequence 2, Appl
43	62	8.0	9880	3	US-08-680-897-1	Sequence 1, Appl
44	61	7.9	456	3	US-09-080-855-30	Sequence 30, Appl
45	61	7.9	2360	1	US-08-466-390-1	Sequence 1, Appl

#### ALIGNMENTS

##### RESULT 1

US-08-669-408B-9  
; Sequence 9, Application US/08669408B  
; Patent No. 6100055

GENERAL INFORMATION:

APPLICANT: GUSS, Bengt  
APPLICANT: JONSSON, Hans  
APPLICANT: LINDBERG, Martin  
APPLICANT: MUELLER, Hans-Peter  
APPLICANT: RANTAMAKI, Liisa K.  
TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING  
PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

##### COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669.408B  
FILING DATE: 03-JUL-1996  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/SE94/00826  
FILING DATE: 06-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 9302855-3  
FILING DATE: 06-SEP-1993  
ATTORNEY/AGENT INFORMATION:

```

Db 548 ACGGCTGGCGATCGTCAGCCAGTTTCGGCGCTCATCGGCATCATCGGCGCTGCTTCA 489
Qy 103 GluGluAlaGluValValGlnSerAsnAlaAlaSerAlaTnpGluLysAlaAla 122
Db 488 AAAGATGAGCGCCACGACAGCGTCAGCAGGAGAGACGCGCCACCGGATGCTCCGTG 429
Qy 123 ThrProIleAlaLeuAspVal 129
Db 428 GCCTCTGTGGCCTCCTGTG 408

```

Search completed: October 13, 2002, 02:25:02  
 Job time : 241.572 secs

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
PS Claim 1: SEQ ID NO 25641; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 3069 BP; 818 A; 832 C; 784 G; 635 T; 0 other;

Alignment Scores:  
Pred. No.: 0.212 Length: 3069  
Score: 81.00 Matches: 26  
Percent Similarity: 43.01% Conservative: 14  
Best Local Similarity: 27.96% Mismatches: 53  
Query Match: 10.42% Indels: 0  
DB: 23 Gaps: 0  
US-09-847-539A-6 (1-159) x AAS89837 (1-3069)  
QY 50 AlaThrThrAlaIleGluAlaAaSerSerAlaLeuGluAlaLeuAlaAaSpGlnThr 69  
DB 157 GCACGAGGTCGCTCCGCGCAAGACATCCGAACGTCACGGAAGCGTCGGAACA 216  
QY 70 AspaLeuGlnSerGluAlaAlaValVallyAlaAaSpnAlaAaSerAspa 89  
DB 217 AGCGCAAGATCCCTCAAAAACGGCTCGCATCGTCAAGCCAGTTTCGGCGGCTCATCGCA 276  
QY 90 LeuGluAlaLeuAlaAaSpGlnThrAspAlaLeuGlnSerGluAlaGluValGln 109  
DB 277 TCATCGCGCTGCTCTCAAAAGATGAGGCGACGACGACGTCAGCGCAAGACGAGC 336  
QY 110 SerAspaAlaAaSerAspaAlaTrpGluLyAlaAlaThrProIleAlaLeuAspVal 129  
DB 337 GCCACGAGGCATCCACGAAGGCGACAGAGGCTGCTGGCAGTCGCGGCGGCGCTCAG 396  
QY 130 LysLysThrLysAspThrLysProValValLysLysGlu 142  
DB 397 AGCAAAAGTACGCGGAATCCGCGGCAACGCGCGCGAG 435

RESULT 40  
AAS74086/C  
ID AAS74086 standard; cDNA; 782 BP.  
XX  
AC AAS74086;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #9890.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.  
OS WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYPSE-) HYSEQ INC.  
XX  
XX Dmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABC09899.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
PS Claim 1: SEQ ID NO 9890; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 782 BP; 135 A; 238 C; 246 G; 163 T; 0 other;  
Alignment Scores:  
Pred. No.: 0.131 Length: 782  
Score: 77.00 Matches: 29  
Percent Similarity: 41.12% Conservative: 15  
Best Local Similarity: 27.10% Mismatches: 63  
Query Match: 9.91% Indels: 0  
DB: 23 Gaps: 0  
US-09-847-539A-6 (1-159) x AAS74086 (1-782)  
QY 23 AsnAlaProGluLysLeuAlaLeuArgAsnGluArgAlaIleAspGluLeuLysLys 42  
DB 728 TCAGCCAGTCCCGCCCGGGAATCGGCAGAAAGCCGCGCATCGTCTGCTTCAACA 669  
QY 43 GlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAaSerAspaLeu 62  
DB 668 GCCACAACGAAGCGTCGCGAAGCCACTGAACAGCCAGCCAGCGAGCTGCTGCTTC 609  
QY 63 GluAlaLeuAlaAaSpGlnThrAspAlaLeuGlnSerGluAlaValVallyAla 82  
DB 608 GCAGCGAAGACATCCGGAACGAGCGGGAAGAGCGTTCGGAACACGAGCAAGTCTCAAAA 549  
QY 83 AspAsnAlaAaSerAspaLeuGluAlaLeuAlaAaSpGlnThrAspAlaLeuGlnSer 102

CC infections, abscesses and spontaneous abortion. This sequence represents  
CC DNA encoding the *S. dysgalactiae* Mig protein of the invention.

XX  
SQ Sequence 2010 BP; 735 A; 393 C; 381 G; 501 T; 0 other;

## Alignment Scores:

Pred. No.: 0.12 Length: 2010  
Score: 81.00 Matches: 17  
Percent Similarity: 62.50% Conservativity: 3  
Best Local Similarity: 53.12% Mismatches: 12  
Query Match: 10.42% Indels: 0  
DB: 24 Gaps: 0

US-09-847-539A-6 (1-159) x AAS18114 (1-2010)

QY 128 AspValLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsnVal 147

DB 1837 GAAGCTAAACACGAGAGCTAAGAAAGAGAGAGCTAGAGAGAGAGCTAAGAAAGCT 1896

QY 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159

DB 1897 GCAACTCTTCTACAACTGGTGAAGGAGCAACCCA 1932

## RESULT 38

AAS77362

ID AAS77362 standard; cDNA; 3069 BP.

XX AC AAS77362;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #13166.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG13175.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX PS Claim 1; SEQ ID No 13166; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94364 represent novel human  
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 3069 BP; 818 A; 832 C; 784 G; 635 T; 0 other;

## Alignment Scores:

Pred. No.: 0.212 Length: 3069  
Score: 81.00 Matches: 26  
Percent Similarity: 43.01% Conservativity: 14  
Best Local Similarity: 27.96% Mismatches: 53  
Query Match: 10.42% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6 (1-159) x AAS77362 (1-3069)

QY 50 AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69

DB 157 GCAGCGAGGTCTGCTTCGCGCAGCAGAGACATCCGAAACGTACGCGAAGCGTCGGAACA 216

QY 70 AspAlaLeuGlnSerGluGluAlaValValLysAlaAspAlaAlaSerAspAla 89

DB 217 ACCGCGAATCTCTCAAAACGGCTGCCGCTCGTCAGCCAGTTCGGCGCGTCATCGGCA 276

QY 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln 109

DB 277 TCATCGGCGTCTGCTTCAAAAGATGAGCGCAGCAGCAGAGCTCAGCAGCAAGAGCAGC 336

QY 110 SerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspVal 129

DB 337 GCCACGACGGCATCCACGAGCGCAGAGGCTGCTGCGAGTTCGCGCGCGCAGCTCAG 396

QY 130 LysLysThrLysAspThrLysProValValLysLysGlu 142

DB 397 AGCAAAAGTACGGCGGAATCCCGCGCAACGCGCGCGAG 435

## RESULT 39

AAS9837

ID AAS9837 standard; cDNA; 3069 BP.

XX AC AAS9837;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #25641.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG25650.

```
Qy 126 AlaLeuAspValLys 130
Db 501 GAAGCGGAGAAAG 515

RESULT 36
AAS82435/c
ID AAS82435 standard; cDNA; 2561 BP.
XX
AC AAS82435;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #18239.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
WPI: 2001-639362/73.
P-PSDB; ABG18248.
XX
New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1: SEQ ID No 18239; 103pp; English.
XX
The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2561 BP; 498 A; 708 C; 666 G; 699 T; 0 other;

Alignment Scores:
Pred. No.: 0.0847 Length: 2561
Score: 83.00 Matches: 26
Percent Similarity: 43.01% Conservative: 14
Best Local Similarity: 27.96% Mismatches: 53
Query Match: 10.68% Indels: 0
DB: 23 Caps: 0
```

```
US-09-847-539A-6 (1-159) x AAS82435 (1-2561)
Qy 50 AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69
Db 1406 GCAGCGAGGTCGTCTCCGCGAGCGAGACATCCGAAACGACGCGAAGCGTCGGAAACA 1347
Qy 70 AspAlaLeuGlnSerGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAla 89
Db 1346 AGCGCAGATCCTCAAAAACGGCTGCCGCATCGTCACGCCAGTTCGGCGCGCTCATCGGA 1287
Qy 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln 109
Db 1286 TCATCGGCGCTGCTTCAAAAGATGAGCGCAGCAGACGAGCGTCCAGCAGCAGCAGC 1227
Qy 110 SerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspVal 129
Db 1226 GCCACGACGGCATCCACGAGGCGAGAGGCTGCTGCGAGTGGCGGCGGACGCTCAG 1167
Qy 130 LysLysThrLysAspThrLysProValValLysLysGlu 142
Db 1166 AGCAAAAGTACGCGGGAATCCGCGGCAACGCGCGCCGAG 1128

RESULT 37
AAS18114
ID AAS18114 standard; DNA; 2010 BP.
XX
AC AAS18114;
XX
DT 26-MAR-2002 (first entry)
XX
DE Streptococcus dysgalactiae Mig DNA.
XX
KW Mig protein; Fc receptor; bacterial infection; Streptococcus agalactiae;
KW streptococcal infection; mastitis; septicemia; meningitis; bacteraemia;
KW impetigo; arthritis; urinary tract infection; abscess; antiinflammatory;
KW spontaneous abortion; antibiotic; antiarthritic; gene therapy; ds.
XX
OS Streptococcus dysgalactiae.
XX
FH Key Location/Qualifiers
FT CDS 1..2010
FT FT /*tag= a
FT FT /product= "S. dysgalactiae Mig protein"
XX
PN WO200196380-A2.
XX
PD 20-DEC-2001.
XX
PF 11-JUN-2001; 2001WO-CA00837.
XX
PR 12-JUN-2000; 2000US-211016P.
XX
PA (UYSA-) UNIV SASKATCHEWAN.
XX
PI Potter RA, Bolton AJ, Song XM;
XX
WPI: 2002-106467/14.
P-PSDB; AAU11350.
XX
Novel vaccine composition comprising Mig protein (Fc receptor) of
PT Streptococcus dysgalactiae useful for treating or preventing
PT streptococcal infection such as mastitis in vertebrates -
XX
PS Example 1; Fig 1; 60pp; English.
XX
The invention relates to a vaccine composition comprising a vehicle and
CC an Fc receptor protein, Streptococcus dysgalactiae Mig protein. The
CC sequence and composition are useful for treating or preventing a
CC bacterial infection e.g. a streptococcal infection which causes mastitis
CC in a vertebrate subject. The vaccine compositions can also be used to
CC treat other streptococcal infections caused by S. agalactiae, such as
CC septicemia, meningitis, bacteraemia, impetigo, arthritis, urinary tract
```







```
Db 3686 AAGGAGCCGCCAGCAGGAGGAGCCCTAGCAAGGTTGTGACTCTGCGGACCCCAAG 3745
Qy 104 GluAlaGluValValGlnSerAspAsuAlaAlaSerAspAlaTrpGluLysAlaAlaThr 123
    |||||
Db 3746 GAGGACAGAGCAATACGGATGAGAGCCAGCCCGCTTCTATTGATGTTCTACCGCC 3805
Qy 124 ProfileAlaLeuaspValLys 131
    |||
Db 3806 CCGCATCAATGATGTCAGCAAG 3829

RESULT 32
AAQ84884
ID AAQ84884 standard; cDNA; 141 BP.
AC
XX
XX
XX 11-OCT-1995 (first entry)
XX
DE Expression cassette for SG-3.
XX
KW Synthetic; protein G; IgG; Fc receptors; heavy chain; leukaemia;
KW lymphoma; cancer; autoimmune disease; ds.
XX
OS Synthetic.
XX
PN WO9506125-A.
XX
PD 02-MAR-1995.
XX
PF 23-AUG-1994; 94WO-US09141.
XX
PR 23-AUG-1993; 93US-0110653.
XX
PA (IMMU-) APPLIED IMMUNE SCI INC.
XX
PI Lee YM, Okarma TB, Talib S;
XX
DR WPI; 1995-106854/14.
DR P-PSDB; AAR71125.
XX
PT New peptide(s) which bind the Fc region of an immunoglobulin -
PT comprising a nontotal portion of the amino acid sequence of
PT Protein A and/or Protein G
XX
PS Claim 20; Fig 2C; 61pp; English.
XX
CC The sequence is that of an expression vector constructed to produce
CC SG-3 (synthetic protein G). SG-3 is capable of specifically binding
CC the constant region of the heavy chain of IgG in the same way as
CC neutral Fc receptors. It can be used to analyse the structure and
CC function of Fc receptors, as well as in antibody production, cell
CC culture, diagnosis and therapy. They can be used to treat disorders
CC such as leukaemia and lymphoma, cancer and immune disorders.
CC See also AAQ84882-7.
XX
SQ Sequence 141 BP; 59 A; 26 C; 29 G; 27 T; 0 other:

Alignment Scores:
Pred. No.: 0.00174 Length: 141
Score: 83.00 Matches: 17
Percent Similarity: 61.29% Conservative: 2
Best Local Similarity: 54.84% Mismatches: 12
Query Match: 10.68% Indels: 0
DB: 16 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ84884 (1-141)

Qy 128 AspValLysThrLysAspThrLysProValLysLysGluArgGlnAsnVal 147
    |||
Db 48 GATACATAAAAGAAGATGCTAAAAAACCAGGCTAAAAAAGAAGATGCTAAAAAAGCT 107
    |||||
Qy 148 AsnThrLeuProThrThrGlyGluSerAsn 158
    |||||

US-09-847-539A-6 (1-159) x AAQ84884 (1-141)

Qy 128 AspValLysThrLysAspThrLysProValLysLysGluArgGlnAsnVal 147
    |||
Db 48 GATACATAAAAGAAGATGCTAAAAAACCAGGCTAAAAAAGAAGATGCTAAAAAAGCT 107
    |||||
Qy 148 AsnThrLeuProThrThrGlyGluSerAsn 158
    |||||
```

```
Db 108 GAAACTCTGCCGACTACTGTGAGGTTCTTAAC 140
RESULT 33
AAQ84887
ID AAQ84887 standard; cDNA; 603 BP.
XX
AC AAQ84887;
XX
XX 11-OCT-1995 (first entry)
XX
DE Chimeric synthetic protein G gene, MD2.
XX
KW Synthetic; protein G; IgG; Fc receptors; heavy chain; leukaemia;
KW lymphoma; cancer; autoimmune disease; ds.
XX
OS Synthetic.
XX
PN WO9506125-A.
XX
PD 02-MAR-1995.
XX
PF 23-AUG-1994; 94WO-US09141.
XX
PR 23-AUG-1993; 93US-0110653.
XX
PA (IMMU-) APPLIED IMMUNE SCI INC.
XX
PI Lee YM, Okarma TB, Talib S;
XX
DR WPI; 1995-106854/14.
DR P-PSDB; AAR71125.
XX
PT New peptide(s) which bind the Fc region of an immunoglobulin -
PT comprising a nontotal portion of the amino acid sequence of
PT Protein A and/or Protein G
XX
PS Claim 20; Fig 4B; 61pp; English.
XX
CC The sequence is that of a synthetic protein G gene, MD-2. The
CC proteins produced by MD-2 are capable of specifically binding
CC the constant region of the heavy chain of IgG in the same way as
CC neutral Fc receptors. They can be used to analyse the structure and
CC function of Fc receptors, as well as in antibody production, cell
CC culture, diagnosis and therapy. They can be used to treat disorders
CC such as leukaemia and lymphoma, cancer and immune disorders.
CC See also AAQ84882-6.
XX
SQ Sequence 603 BP; 218 A; 120 C; 126 G; 139 T; 0 other:

Alignment Scores:
Pred. No.: 0.0122 Length: 603
Score: 83.00 Matches: 17
Percent Similarity: 61.29% Conservative: 2
Best Local Similarity: 54.84% Mismatches: 12
Query Match: 10.68% Indels: 0
DB: 16 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ84887 (1-603)

Qy 128 AspValLysThrLysAspThrLysProValLysLysGluArgGlnAsnVal 147
    |||
Db 48 GATACATAAAAGAAGATGCTAAAAAACCAGGCTAAAAAAGAAGATGCTAAAAAAGCT 546
    |||||
Qy 148 AsnThrLeuProThrThrGlyGluSerAsn 158
    |||||
Db 547 GAAACTCTGCCGACTACTGTGAGGTTCTTAAC 579

RESULT 34
AAQ84886
ID AAQ84886 standard; cDNA; 627 BP.
XX
AC AAQ84886;
XX
```



CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS84197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 3411 BP; 921 A; 898 C; 918 G; 674 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.0451 Length: 3411  
 Score: 86.00 Matches: 26  
 Percent Similarity: 43.30% Conservative: 16  
 Best Local Similarity: 26.80% Mismatches: 55  
 Query Match: 11.07% Indels: 0  
 DB: 23 Gaps: 0

US-09-847-539A-6 (1-159) x AAS82434 (1-3411)

Qy 46 GluAspLysGluAlaThrThrAlaIleGluAlaSerSerAspAlaLeuGluAlaLeu 65  
 Db 2026 CAGTCCAGCGCAGCGAGTCTGCTTCCGACGAGACATCCGAACGACGCGAAA 2085  
 Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluAlaValValLysAlaAspAla 85  
 Db 2086 CGCTCGGAACACGCGAGATCTCTAAACGCGTCGCCGATCGTACGCCAGTCCGCG 2145  
 Qy 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAla 105  
 Db 2146 CGGTCATCGGCATCATCGCGCTGCTTCARAGATGAGGCGACACAGCGTCAGCA 2205  
 Qy 106 GluValValGlnSerAspAlaAlaSerAspAlaTrpGluLysAlaAlaThrPro 125  
 Db 2206 GCGAAGACGACGCGCACGACGCGATCCACGAAGCGACAGAGGCTGCTGGCAGTCCGACG 2265  
 Qy 126 AlaLeuAspValLysLysThrLysAspThrLysProValValLysLysGlu 142  
 Db 2266 GCGGAGCTCAGACGAAGTACGCGGGAATCCGCGGCAACGCGCGCGAG 2316

RESULT 29  
 AAQ89196  
 ID AAQ89196 standard; DNA; 2526 BP.

XX AAQ89196;

XX 22-SEP-1995 (first entry)  
 XX S. dysgalactiae mig gene.  
 XX MIG; fast alpha-2-macroglobulin binding protein; FAM;  
 KW plasma proteinase-inhibitor binding protein; ss.  
 XX Streptococcus dysgalactiae strain SCL.

PH key Location/Qualifiers  
 FT promoter 394..399  
 FT /\*tag= a  
 FT /note= "putative promoter region"  
 FT promoter 416..421  
 FT /\*tag= b  
 FT /note= "putative promoter region"  
 FT promoter 447..453  
 FT /\*tag= c  
 FT /note= "putative promoter region"  
 FT promoter 467..472  
 FT /\*tag= d  
 FT /note= "putative promoter region"  
 FT RBS 494..499

FT CDS /\*tag= e  
 FT 506..2500  
 FT /\*tag= f  
 FT 506..595  
 FT /\*tag= g  
 FT 596..2497  
 FT /\*tag= h  
 XX  
 PN W09507296-A.  
 XX 16-MAR-1995.  
 XX 06-SEP-1994; 94WO-SE00826.  
 XX 06-SEP-1993; 93SE-0002855.  
 XX (GUSS/) GUSS B.  
 XX (JONS/) JONSSON H.  
 XX (LIND/) LINDBERG M.  
 XX (MOEL/) MOELLER H.  
 XX (RANT/) RANTAMAKI L K.  
 XX Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;  
 XX WPI; 1995-123382/16.  
 XX P-PSDB; AAR71929.  
 XX DNA encoding fast alpha 2-macroglobulin-binding proteins - used  
 XX to obtain prods for sepn., detection or quantification or for  
 XX binding inhibition  
 XX Disclosure; Page 31; 50pp; English.  
 XX A phage lambda GEM-11 library of S. dysgalactiae SCL DNA was analyzed  
 XX for fast-acting alpha-2-macroglobulin (FAM)- and IgG-binding  
 XX activity. One clone, which expressed both activities, was  
 XX analyzed to obtain DNA encoding the FAM-binding protein, MIG  
 XX (AAR71929). The mig gene is given in AAQ89196.  
 XX SQ Sequence 2526 BP; 924 A; 449 C; 470 G; 683 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.0423 Length: 2526  
 Score: 85.00 Matches: 18  
 Percent Similarity: 62.50% Conservative: 2  
 Best Local Similarity: 56.25% Mismatches: 12  
 Query Match: 10.94% Indels: 0  
 DB: 16 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ89196 (1-2526)

Qy 128 AspValLysLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsnVal 147  
 Db 2327 GACGCTAAAACCCAGAACGCTAAGAAAGAAAGAGCTAAGAAAGAAAGCTAAGAAAGCT 2386  
 Qy 148 AsnThrLeuProThrThrGlyGluSerAsnPro 159  
 Db 2387 GCAACTCTCTCTACAACTGGTGAAGGAAGCAACCCA 2422

RESULT 30  
 ABL03677  
 ID ABL03677 standard; cDNA; 2129 BP.  
 XX  
 AC ABL03677;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5513.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.



```
FT      /tag= e
FT      /label= B2
FT      459..530
FT      /tag= f
FT      /label= A3
FT      531..566
FT      /tag= g
FT      /label= S
FT      567..731
FT      /tag= h
FT      /label= C1
FT      732..776
FT      /tag= i
FT      /label= D1
FT      777..941
FT      /tag= j
FT      /label= C2
FT      942..986
FT      /tag= k
FT      /label= D2
FT      987..1151
FT      /tag= l
FT      /label= C3
FT      1152..1361
FT      /tag= m
FT      /label= W
FT      1362..1469
FT      /tag= n
FT      /label= M
XX
XX      WO8705631-A.
XX
XX      24-SEP-1987.
XX
XX      20-MAR-1987; 87WO-SE00145.
XX
XX      21-MAR-1986; 86SE-0001325.
XX
XX      (PHAA ) PHARMACIA AB.
XX      (GUSS/) GUSS B M.
XX
XX      Guss BM, Lindberg KM, Flock JI, Uhlen CEM;
XX
XX      WPI; 1987-277686/39.
XX      P-PSDB; AAP70468.
XX
XX      New recombinant DNA molecules - for producing proteins with
XX      Igg-binding specificity of protein G or proteins A and G
XX      Example; Fig 2; 39pp; English.
XX
XX      A recombinant DNA molecule containing a nucleotide sequence which
XX      codes for a protein or polypeptide having the same Igg specificity as
XX      protein G from Streptococcus G148 (AAN70757) is claimed. See, for
XX      example, AAN70754, AAN70755 and AAN70756.
XX
XX      Sequence 1469 BP; 545 A; 280 C; 297 G; 347 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.: 0.00377 Length: 1469
XX      Score: 90.00 Matches: 18
XX      Percent Similarity: 62.50% Conservative: 2
XX      Best Local Similarity: 56.25% Mismatches: 12
XX      Query Match: 11.58% Indels: 0
XX      DB: 8 Gaps: 0
XX
XX      US-09-847-539A-6 (1-159) x AAN70757 (1-1469)
XX
XX      Qy 128 AspValLysThrLysAspThrLysProValValLysGluGluArgGlnAsnVal 147
XX      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      Db 1270 GATACTAAGAAGACAGACTCTAAATAAACCCAGAGCTAAGAAAGACGCTAAGAAAGCT 1329
XX
XX      Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
```

```
Db 1330 CAAACTCTTCTACAACTGGTGAAGGCAGCAACCCA 1365
RESULT 26
ABL20337
ID ABL20337 standard; DNA; 2121 BP.
XX
XX ABL20337;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 12484.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX      pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX      11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX      genes from Drosophila and for elucidating cell signalling and cell-cell
XX      interactions -
XX
XX Claim 1; SEQ ID NO 12484; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX      capable of detecting 1000 or more genes from Drosophila. The invention is
XX      useful in developmental biology and in elucidating cell signalling and
XX      cell-cell interactions in higher eukaryotes for the development of
XX      insecticides, therapeutics and pharmaceutical drugs. The invention
XX      discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX      sequences (ABL01840-ABL16175) and the encoded proteins
XX      (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2121 BP; 533 A; 638 C; 587 G; 363 T; 0 other;
XX
XX Alignment Scores:
XX      Pred. No.: 0.00617 Length: 2121
XX      Score: 90.00 Matches: 30
XX      Percent Similarity: 45.10% Conservative: 16
XX      Best Local Similarity: 29.41% Mismatches: 56
XX      Query Match: 11.58% Indels: 0
XX      DB: 23 Gaps: 0
XX
XX US-09-847-539A-6 (1-159) x ABL20337 (1-2121)
XX
XX Qy 26 GluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLysLysGlnAlaIle 45
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      Db 1342 GAAAGCAGGCTGCTGCTAAGCCCAAAAGGCTGCCGAGGAGGCCGCCCAAAATCGCTGCC 1401
XX
XX Qy 46 GluAspLysGluAlaThrThrAlaIleGluAlaIleSerAspAlaLeuGluAlaLeu 65
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      Db 1402 GAAGAGGCTCTCTGGCTGAGCTGCCGCCCAAAAGCAGCGGAGAGCTAAAGCCCTA 1461
XX
XX Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleValLysAlaAspAsnAla 85
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

PS Disclosure; Page 29-30; 38pp; English.
XX A phage lambda EMBL3 library of S. dysgalactiae 8215 DNA was
CC analyzed for alpha-2-macroglobulin-, BSA- and IgG-binding activity.
CC Clone lambda-SD1, which expressed all 3 activities, was analyzed to
CC obtain DNA encoding the SA-binding protein, MAG. The mag gene
CC (given in Q86080) encoded a 44 kDa MAG protein (R71670).
CC Recombinant, immobilized MAG was used for serum albumin affinity
CC purification, detection and assay.
XX Sequence 1555 BP; 582 A; 242 C; 305 G; 426 T; 0 other;

Alignment Scores:
Pred. No.: 0.0029 Length: 1555
Score: 91.00 Matches: 19
Percent Similarity: 62.50% Conservative: 1
Best Local Similarity: 59.38% Mismatches: 12
Query Match: 11.71% Indels: 0
DB: 16 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ86080 (1-1555)
QY 128 AspValLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1356 GATACTAAGAAAGAGATGCTAAACCAAGAGCTAAGAAAGAGCTAAGAAAGCT 1415

QY 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1416 GCAACTCTTCTCACTGCTGAAGGAGCAACCCA 1451

RESULT 24
AAQ89197
ID AAQ89197 standard; DNA; 1555 BP.
XX
XX AAQ89197;
XX
XX 22-SEP-1995 (first entry)
XX
XX S. dysgalactiae mag gene.
XX
XX Mag; fast alpha-2-macroglobulin binding protein; FAM;
KW plasma proteinase-inhibitor binding protein; ss.
XX
XX Streptococcus dysgalactiae strain 8215.
XX
XX Key Location/Qualifiers
FH promoter 198..203
FT /*tag= a
FT /*note= "putative promoter region"
FT 230..235
FT /*tag= b
FT /*note= "putative promoter region"
FT 249..254
FT /*tag= c
FT /*note= "putative promoter region"
FT 264..269
FT /*tag= d
FT /*note= "putative promoter region"
FT 276..281
FT RBS /*tag= e
FT 288..1527
FT CDS /*tag= f
FT sig_peptide 288..389
FT /*tag= g
FT mat_peptide 390..1524
FT /*tag= h
XX
XX WO9507296-A.
XX
XX 16-MAR-1995.
XX
XX 06-SEP-1994; 94WO-SE00826.
XX

```

```

PR 06-SEP-1993; 93SE-0002855.
XX
XX (GUSS/) GUSS B.
PA (JONS/) JONSSON H.
PA (LIND/) LINDBERG M.
PA (MUEL/) MUELLER H.
PA (RANT/) RANTAMAKI L K.
XX
PI Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;
XX WPI; 1995-123382/16.
DR P-PSDB; AAR71928.
XX
XX DNA encoding fast alpha 2-macro:globulin-binding proteins - used
PT to obtain prods. for seph., detection or quantification or for
PT binding inhibition
XX
XX Disclosure; Fig.4; 50pp; English.
XX
XX A phage lambda EMBL library of S. dysgalactiae 8215 DNA was analyzed
CC for fast-acting alpha-2-macroglobulin (FAM)-, BSA- and IgG-binding
CC activity. Clone lambda-SD1, which expressed all 3 activities, was
CC analyzed to obtain DNA encoding the FAM-binding protein, MAG. The
CC mag gene (given in Q89197) encodes a 44 kDa protein (R71929).
XX
XX Sequence 1555 BP; 583 A; 242 C; 304 G; 426 T; 0 other;

Alignment Scores:
Pred. No.: 0.0029 Length: 1555
Score: 91.00 Matches: 19
Percent Similarity: 62.50% Conservative: 1
Best Local Similarity: 59.38% Mismatches: 12
Query Match: 11.71% Indels: 0
DB: 16 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ89197 (1-1555)
QY 128 AspValLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1356 GATACTAAGAAAGAGATGCTAAACCAAGAGCTAAGAAAGAGCTAAGAAAGCT 1415

QY 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1416 GCAACTCTTCTCACTGCTGAAGGAGCAACCCA 1451

RESULT 25
AAQ89197
ID AAQ89197 standard; DNA; 1469 BP.
XX
XX AAQ89197;
XX
XX 21-MAY-1991 (first entry)
XX
XX Sequence encoding polypeptide possessing IgG-binding activity of
DE protein G from Streptococcus G148.
XX
XX Antibody-binding; IgG; IgA; immunoglobulin; ss.
XX
XX Streptococcus G148.
XX
XX Key Location/Qualifiers
FH CDS 1..1443
FT /*tag= a
FT misc_feature 9..80
FT /*tag= b
FT /*label= A1
FT misc_feature 81..233
FT /*tag= c
FT /*label= B1
FT misc_feature 234..305
FT /*tag= d
FT /*label= A2
FT misc_feature 306..458

```



```

RESULT 22
AAS80225
ID AAS80225 standard; cDNA; 1383 BP.
XX
AC AAS80225;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #16029.
XX
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX P-PSDB; ABG16038.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 1; SEQ ID NO 16029; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS84197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1383 BP; 331 A; 349 C; 439 G; 264 T; 0 other;
XX

Alignment Scores:
Pred. No.: 0.00126 Length: 1383
Score: 93.00 Matches: 24
Percent Similarity: 54.41% Conservative: 13
Best Local Similarity: 35.29% Mismatches: 31
Query Match: 11.97% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6 (1-159) x AAS80225 (1-1383)

QY 50 AlathrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69

```

```

Db 1162 GCAGAGTCCTCAAAAACGCGCGCCACCAAGTCCGCGTCCGCGGAAACGTCAGAAACG 1221
XX
QY 70 AspAlaLeuGlnSerGluGluAlaAlaValValValValAlaAspAsnAlaAlaSerAspAla 89
XX
Db 1222 AATGCTGCACGCTCACAACAATCAGCGCCACGCTCTGCTCCACCGGCGCAGAAAGCG 1281
XX
QY 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln 109
XX
Db 1282 TCAGAGGCCGCCACTTCAGCACGAGATGCGGTGCCTCAAAAGAGCGCAGCAAAATCATCA 1341
XX
QY 110 SerAspAsnAlaAlaSerAspAla 117
XX
Db 1342 GAAACGAAACGATCATCAAGTGCC 1365
XX
RESULT 23
AAO86080
ID AAO86080 standard; DNA; 1555 BP.
XX
AC AAO86080;
XX
DT 21-SEP-1995 (first entry)
XX
DE S. dysgalactiae mag gene.
XX
KW Mag gene; serum albumin binding protein; protein stabilization;
XX vaccine; ss.
XX
OS Streptococcus dysgalactiae strain 8215.
XX
XX Key Location/Qualifiers
FH promoter 198..203
FT /*tag= a
FT /*note= "putative promoter sequence"
FT promoter 230..235
FT /*tag= b
FT /*note= "putative promoter sequence"
FT promoter 249..254
FT /*tag= c
FT /*note= "putative promoter sequence"
FT promoter 264..269
FT /*tag= d
FT /*note= "putative promoter sequence"
FT RBS 276..281
FT /*tag= e
FT CDS 288..1527
FT /*tag= f
FT sig_peptide 288..389
FT /*tag= g
FT mat_peptide 390..1524
FT /*tag= h
XX
XX WO9507300-A.
XX
XX 16-MAR-1995.
XX
XX 06-SEP-1994; 94WO-SB00825.
XX
XX 06-SEP-1993; 93SE-0002856.
XX
XX (GUSS/) GUSS B.
XX (JONS/) JONSSON H.
XX (LIND/) LINDBERG M.
XX
XX Guss B, Jonsson H, Lindberg M;
XX
XX WPI: 1995-123386/16.
XX P-PSDB; AAR71670.
XX
XX DNA encoding a serum albumin binding protein - used to obtain
XX prods. for seps., detection, quantification, protein
XX stabilisation or vaccine development
XX

```



DB 160 TCAGAGACAACACTGGCTTTTCGTAATGAAGAGAGTGCT 198  
 ::: ||||||||||||||||||| |||

## RESULT 18

AAQ06009

ID AAQ06009 standard; DNA: 798 BP.

XX AC AAQ06009;

XX 17-JAN-1991 (first entry)

XX Sequence encoding Protein G variant.

XX Immunoglobulin; ds.

XX Streptococcus sp. Lancefield Group G strain.

XX US4956296-A.

XX 11-SEP-1990.

XX 20-JUN-1988; 88US-0209236.

XX 20-JUN-1988; 88US-0209236.

XX 14-FEB-1986; 86US-0829354.

XX 23-APR-1986; 86US-0854887.

XX 17-FEB-1987; 87WO-US00329.

XX 19-JUN-1987; 87US-0063959.

XX (GENE-) GENEX CORP.

XX Fahnestock SR;

XX WPI: 1990-297491/39.

XX P-PSDB; AAR07004.

XX Recombinant Protein G variants - obt'd. using a cloned gene

XX encoding Protein G from Streptococcus sp., used for binding

XX immunoglobulin.

XX Claim 4; Column 47; 48pp; English.

XX Sequence may be incorporated into a non-pathogenic host eg. E.coli,

XX where they may be expressed at high levels. The proteins have a

XX higher binding efficiency and capacity for immunoglobulin, and may

XX be used for purifying, detecting and isolating antibodies.

XX Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

SQ

Alignment Scores:

Pred. No.: 1.04e-05 Length: 798

Score: 105.00 Matches: 23

Percent Similarity: 72.73% Conservative: 1

Best Local Similarity: 69.70% Mismatches: 9

Query Match: 13.51% Indels: 0

DB: 11 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ06009 (1-798)

QY 4 ProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuGlyAsn 23

DB 100 CCAATCGAAGATACCCCAATTTTCGTAATGGTGGTCAATTAACATACTTCTGGGAAT 159

QY 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAla 36

DB 160 TCAGAGACAACACTGGCTTTTCGTAATGAAGAGAGTGCT 198

RESULT 19

AAQ10007

ID AAQ10007 standard; DNA: 798 BP.

XX AC AAQ10007;

XX 17-JAN-1991 (first entry)

XX Sequence encoding Protein G variant.

XX Immunoglobulin; ds.

XX Streptococcus sp. Lancefield Group G strain.

XX US4956296-A.

XX 11-SEP-1990.

XX 20-JUN-1988; 88US-0209236.

XX 20-JUN-1988; 88US-0209236.

XX 14-FEB-1986; 86US-0829354.

XX 23-APR-1986; 86US-0854887.

XX 17-FEB-1987; 87WO-US00329.

XX 19-JUN-1987; 87US-0063959.

XX (GENE-) GENEX CORP.

XX Fahnestock SR;

XX WPI: 1990-297491/39.

XX P-PSDB; AAR07004.

XX Recombinant Protein G variants - obt'd. using a cloned gene

XX encoding Protein G from Streptococcus sp., used for binding

XX immunoglobulin.

XX Claim 4; Column 47; 48pp; English.

XX Sequence may be incorporated into a non-pathogenic host eg. E.coli,

XX where they may be expressed at high levels. The proteins have a

XX higher binding efficiency and capacity for immunoglobulin, and may

XX be used for purifying, detecting and isolating antibodies.

XX Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

SQ

DT 13-MAR-1991 (first entry)

XX Type 4 CX7809 protein G variant.

XX Immunoglobulins; Ig; ds.

XX Streptococcus sp CX7809.

XX US4977247-A.

XX 11-DEC-1990.

XX 19-MAY-1989; 89US-0354264.

XX 19-MAY-1989; 89US-0354264.

XX 14-FEB-1986; 86US-0829354.

XX 23-APR-1986; 86US-0854887.

XX 17-FEB-1987; 87WO-US00329.

XX 19-JUN-1987; 87US-0063959.

XX 20-JUN-1988; 88US-0209236.

XX (GENE-) GENEX CORP.

XX Fahnestock SR, Lee T, Wroble MH;

XX WPI: 1991-006758/01.

XX P-PSDB; AAR10011.

XX Immobilised protein G variants - used for detection, isolation

XX and purificn. immunoglobulin(s) and immunoglobulin fragments

XX Disclosure; Column 11; 52pp; English.

XX Protein G variant product carries active regions B1 and B2, it may

XX be immobilised and exhibits different binding profiles. The bound

XX protein is useful in purification and detection of Igs and fragments.

XX Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

SQ

Alignment Scores:

Pred. No.: 1.04e-05 Length: 798

Score: 105.00 Matches: 23

Percent Similarity: 72.73% Conservative: 1

Best Local Similarity: 69.70% Mismatches: 9

Query Match: 13.51% Indels: 0

DB: 12 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ10007 (1-798)

QY 4 ProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuGlyAsn 23

DB 100 CCAATCGAAGATACCCCAATTTTCGTAATGGTGGTCAATTAACATACTTCTGGGAAT 159

QY 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAla 36

DB 160 TCAGAGACAACACTGGCTTTTCGTAATGAAGAGAGTGCT 198

RESULT 20

AAQ64648

ID AAQ64648 standard; DNA: 798 BP.

XX AC AAQ64648;

XX 06-JAN-1995 (first entry)

XX IgG-binding Streptococcus Protein G variant coding sequence.

XX Streptococcus Protein G; variant; IgG binding activity;

XX immunoglobulin; Lancefield Group G; bacterial Fc receptor; ss.

XX Streptococcus sp.

XX Key Location/Qualifiers

FH

KW liposomal delivery complex; connecting moiety; antibody; Fc region;  
 KW diagnosis; therapeutic agent; antibiotic; antidepressant; antiviral;  
 KW antitumorigenic; cytokine; hormone; imaging agent; neurotransmitter;  
 KW stimulant; cytotoxic agent; malignant cell; ss.

XX Streptococcus sp.

XX Key Location/Qualifiers

FT CDS 292..603

FT FT /\*tag= a

FT FT 705..1262

FT FT /\*tag= b

FT FT /product= "protein G"

FT FT /transl\_except= (pos:705..707,aa:Met)

PN WO9959643-A2.

XX 25-NOV-1999.

XX 19-MAY-1999; 99WO-US11177.

XX 20-MAY-1998; 98US-0086347.

XX (SDGS-) SDG INC.

XX Lau JR;

XX WPI: 2000-062383/05.

XX P-PSDB; AAY57610, AAY57611.

XX New liposome constructs comprising a liposome connected to an antibody,  
 used, e.g. for delivery of cytotoxic agents to malignant cells -

XX Disclosure; Fig 1; 32pp; English.

XX The present invention describes a construct for connecting an antibody  
 or antibody fragment to a liposome. The construct comprises protein G'  
 and a linking moiety for connecting the protein G' to the liposome.  
 XX Also described are: (1) a liposomal delivery vehicle comprising: (a) a  
 liposome; and (b) a connecting moiety connected to the liposome, which  
 specifically binds the Fc region of an antibody, for connecting the  
 antibody to the liposome; and (2) forming a liposomal carrier comprising:  
 XX (a) providing liposomal components having binding moieties, for forming  
 a liposome; (b) providing a construct comprising a linking moiety and a  
 connecting moiety bound together; and (c) combining the liposomal  
 components and the construct and sonicating the combination so that the  
 binding moieties are exposed on a surface of the liposome, so that the  
 the linking moiety, and to facilitate the binding between them. The  
 products can be used for the delivery of diagnostic or therapeutic  
 agents. The liposomes may contain or may be associated with a diagnostic  
 or therapeutic agent, e.g. antibiotics, antidepressants,  
 XX antitumorigenics, antivirals, cytokines, hormones, imaging agents,  
 neurotransmitters, or stimulants. They can be used particularly for the  
 delivery of cytotoxic agents to malignant cells. The protein G'  
 connecting moiety provides a liposomal delivery complex having improved  
 targeting efficiency. As a result of the binding between protein G' and  
 the Fc region of antibodies, protein G' shields the Fc regions of the  
 attached antibodies from non-specific binding to cell-surfaces, other  
 proteins, and anatomical structures. The present sequence encodes  
 XX an unidentified protein and protein G'.

XX Sequence 1576 BP; 572 A; 270 C; 330 G; 404 T; 0 other;

XX Alignment Scores:

Pred. No.:	9.57e-11	Length:	1576
Score:	142.00	Matches:	37
Percent Similarity:	60.61%	Conservative:	3
Best Local Similarity:	56.06%	Mismatches:	26
Query Match:	18.28%	Indels:	1
DB:	21	Gaps:	0

US-09-847-539A-6 (1-159) x AA247931 (1-1576)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 |||||  
 Db 391 GTTGACTCACCATTGGAAGATACCCCAATTATTCGTAATGGTGGTGAATTAATCTT 450  
 |||||  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
 |||||  
 Db 451 CTGGGAATTCAGAGACAACACTGGCTTTCGCTAATGAAGAGAGTCTACAGCTGATTG 510  
 |||||  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaIleAspSerAsp 60  
 |||||  
 Db 511 ACAGCAGCAGCGGTAGCCGATACTGTGGCAGCAGCGGAGCTGAAATGCTGGGAGCA- 569  
 |||||

QY 61 AlaLeuGluAlaLeuAla 66  
 |||||

Db 570 GCTTGGGAAGCAGCGGCA 587

RESULT 17

AAAN91099

ID AAAN91099 standard; DNA; 798 BP.

XX

AC AAAN91099;

XX 04-JUL-1990 (first entry)

XX Protein G gene variant.

XX Protein G; immunoglobulin; Fc receptor; ds.

XX Streptococcus sp.

XX WO8810306-A.

XX 29-DEC-1988.

XX 20-JUN-1988; 88WO-US02084.

XX 19-JUN-1987; 87US-0063959.

XX (GENE-) GENEX CORP.

XX Fahnstock SR;

XX WPI: 1989-023848/03.

XX P-PSDB; AAP94785.

XX Cloned protein G variant genes -

XX expressing proteins having immunoglobulin-binding properties of  
 protein G and derived from Streptococcus sp.  
 XX Claim 4; Page 83; 116pp; English.

XX Gene encodes protein G variant of non-pathogenic streptococcus allowing  
 isolation of the protein and variants, useful as bacterial Fc receptors  
 eg in purification and detection of Abs., screening of hybridoma clones  
 and treatment of disease.

XX Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

XX Alignment Scores:

Pred. No.:	1.04e-05	Length:	798
Score:	105.00	Matches:	23
Percent Similarity:	72.73%	Conservative:	1
Best Local Similarity:	69.70%	Mismatches:	9
Query Match:	13.51%	Indels:	0
DB:	10	Gaps:	0

US-09-847-539A-6 (1-159) x AAAN91099 (1-798)

QY 4 ProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuGlyAsn 23  
 |||||

Db 100 CCAATCGAAGATACCCCAATTATTCGTAATGGTGAATTAATCTTCTGGGAAAT 159  
 |||||

QY 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAla 36

```

XX (GENE-) GENEX CORP.
XX PI Fahnstoc SR, Lee T, Wroble MH;
XX DR WPI; 1991-006758/01.
XX DR P-PSDB; AAR10005.
XX
XX Immobilised protein G variants - used for detection, isolation
XX and purificn. immunoglobulin(s) and immunoglobulin fragments
XX
XX Disclosure; Fig 9; 52pp; English.
XX
XX Protein G gene product may be modified allowing the variant to
XX be immobilised and exhibit different binding profiles. The bound
XX protein is useful in purification and detection of Igs and fragments.
XX
XX Sequence 2383 BP; 867 A; 409 C; 481 G; 626 T; 0 other;

Alignment Scores:
Pred. No.:      2.05e-12      Length:      2383
Score:          155.00      Matches:      37
Percent Similarity: 62.12%      Conservative: 4
Best Local Similarity: 56.06%      Mismatches: 25
Query Match:     19.95%      Indels:      0
DB:              12          Gaps:          0

US-09-847-539A-6 (1-159) x AAQ10002 (1-2383)
QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
DB 676 GTTGACTCACCACAGATACCCCAATATTGTAATGGTGGTGAATTAACATAATCTT 735
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
DB 736 CTGGGAATTCAGAGACAACACTGGCTTTGCGTAAATGAAGAGAGTCTACAGCTGATTG 795
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
DB 796 ACAGCAGCAGCGGTAGCCGATACTGTGGCAGCAGCGCAGCTGCTGCTGCTGCTGCTG 855
QY 61 AlaLeuGluAlaLeuAla 66
DB 856 GCTTGGGAAGCAGCGGCA 873

RESULT 15
AAQ75036
ID AAQ75036 standard; DNA; 2383 BP.
XX
XX AC AAQ75036;
XX
XX 10-JAN-1995 (first entry)
XX
XX Streptococcus Protein G gene derived from strain GX7805.
XX
XX Streptococcus Protein G; variant; IgG binding activity;
XX immunoglobulin; Lancefield Group G; bacterial Fc receptor; ss.
XX
XX Streptococcus sp. GX7805.
XX
XX Key Location/Qualifiers
XX -35_signal 469
XX /*tag= a
XX -10_signal 486..491
XX /*tag= b
XX RBS 558..564
XX /*tag= c
XX CDS 577..2358
XX /*tag= d
XX /product= Protein_G
XX /transl_except= pos:1381..1383, aa:Gln
XX /transl_except= pos:1702..1704, aa:Leu
XX /transl_except= pos:1972..1974, aa:Gly

```

```

FT XX /transl_except= pos:2350..2352, aa:Gly
XX
XX US5312901-A.
XX
XX 17-MAY-1994.
XX
XX 14-FEB-1986; 86US-0829354.
XX
XX 14-FEB-1986; 86US-0829354.
XX
XX 23-APR-1986; 86US-0854887.
XX
XX 19-JUN-1987; 87US-0063959.
XX
XX 20-JUN-1988; 88US-0203236.
XX
XX 19-JUN-1990; 90US-0540169.
XX
XX 21-APR-1992; 92US-0871539.
XX
XX (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX
XX Fahnstoc SR;
XX
XX WPI; 1994-159179/19.
XX
XX P-PSDB; AAR62944.
XX
XX New recombinant streptococcal protein G variants - useful for
XX antibody detection and purification and for therapy
XX
XX Example 5; Fig 9; 48pp; English.
XX
XX A 2.4kb HindIII fragment containing the entire coding sequence for
XX Protein G was isolated from Streptococcus GX7805 using the 1.9kb
XX Protein G coding sequence from Streptococcus GX7805. The protein G
XX has IgG-binding activity which has been localised to the B repeating
XX structure. Streptococcal Protein G variants comprising the B domains
XX are claimed.
XX
XX Sequence 2383 BP; 867 A; 409 C; 482 G; 625 T; 0 other;

Alignment Scores:
Pred. No.:      2.05e-12      Length:      2383
Score:          155.00      Matches:      37
Percent Similarity: 62.12%      Conservative: 4
Best Local Similarity: 56.06%      Mismatches: 25
Query Match:     19.95%      Indels:      0
DB:              15          Gaps:          0

US-09-847-539A-6 (1-159) x AAQ75036 (1-2383)
QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
DB 676 GTTGACTCACCACAGATACCCCAATATTGTAATGGTGGTGAATTAACATAATCTT 735
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
DB 736 CTGGGAATTCAGAGACAACACTGGCTTTGCGTAAATGAAGAGAGTCTACAGCTGATTG 795
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
DB 796 ACAGCAGCAGCGGTAGCCGATACTGTGGCAGCAGCGCAGCTGCTGCTGCTGCTGCTG 855
QY 61 AlaLeuGluAlaLeuAla 66
DB 856 GCTTGGGAAGCAGCGGCA 873

RESULT 16
AAZ47931
ID AAZ47931 standard; cDNA; 1576 BP.
XX
XX AC AAZ47931;
XX
XX 10-MAR-2000 (first entry)
XX
XX Streptococcus strain G 148 protein encoding cDNA.
XX
XX Streptococcus strain G 148; protein G'; protein G primer; liposome;

```

PT antibody detection and purification and for therapy  
 XX Example 2; Fig 3; 48pp; English.  
 XX A 1.9kb HindIII fragment containing the entire coding sequence for  
 CC Protein G was isolated from Streptococcus GX7809. The protein G has  
 CC IgG-binding activity which has been localised to the B repeating  
 CC structure. Streptococcal Protein G variants comprising the B domains  
 CC are claimed.  
 XX Sequence 1950 BP; 706 A; 323 C; 398 G; 523 T; 0 other;

Alignment Scores:  
 Pred. No.: 1.57e-12 Length: 1950  
 Score: 155.00 Matches: 37  
 Percent Similarity: 62.12% Conservative: 4  
 Best Local Similarity: 56.06% Mismatches: 25  
 Query Match: 19.95% Indels: 0  
 DB: 15 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ064644 (1-1950)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 DB 677 GTTGATTCCACCAATCGAAGATACCCCAATTATTCGTAATGGTGGTGAATTAACCTAATCTT 736  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluArgAlaIleAspGluLeu 40  
 DB 737 CTGGGGAAATTCAGACACACACGCTGGCTTGCCTAATGAAGAGAGTGCTACAGCTGATTG 796  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60  
 DB 797 ACAGCAGCAGGGTAGCCGATCTGTGCGACGACGCGCAGCTGAAATGCTGGGGCAGCA 856  
 QY 61 AlaLeuGluAlaLeuAla 66  
 DB 857 GCTTGGGAAGCAGCGGCA 874

#### RESULT 13

AAQ06019  
 ID AAQ06019 standard; DNA; 2383 BP.

XX AC AAQ06019;

DT 17-JAN-1991 (first entry)

DE Sequence encoding Protein G with three active sites.

XX Immunoglobulin; ds.

OS Streptococcus sp. Lancefield Group G strain.

XX Key Location/Qualifiers

FT CDS 1..2355

FT /\*tag= a

PN US4956296-A.

XX 11-SEP-1990.

XX 20-JUN-1988; 88US-0209236.

XX 20-JUN-1988; 88US-0209236.

PR 14-FEB-1986; 86US-0829354.

PR 23-APR-1986; 86US-0854887.

PR 17-FEB-1987; 87WO-US00329.

PR 19-JUN-1987; 87US-0063959.

XX (GENE-) GENEX CORP.

XX Fahnestock SR;

XX WPI; 1990-297491/39.

DR P-PSDB; AAR07014.  
 XX Recombinant Protein G variants - Obt'd. using a cloned gene  
 PT encoding Protein G from Streptococcus sp., used for binding  
 PT immunoglobulin.  
 XX Disclosure; Fig 9; 48pp; English.  
 XX Fragments and variants of the sequence are claimed esp. where  
 CC incorporated into a non-pathogenic host eg. E.coli, and expressed  
 CC at high levels.  
 CC The variants have a higher binding efficiency and capacity for  
 CC immunoglobulin, and may be used for purifying, detecting and  
 CC isolating antibodies.  
 XX Sequence 2383 BP; 867 A; 411 C; 479 G; 626 T; 0 other;

Alignment Scores:  
 Pred. No.: 2.05e-12 Length: 2383  
 Score: 155.00 Matches: 37  
 Percent Similarity: 62.12% Conservative: 4  
 Best Local Similarity: 56.06% Mismatches: 25  
 Query Match: 19.95% Indels: 0  
 DB: 11 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ06019 (1-2383)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 DB 676 GTTGACTCACCAATCGAAGATACCCCAATTATTCGTAATGGTGGTGAATTAACCTAATCTT 735  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluArgAlaIleAspGluLeu 40  
 DB 736 CTGGGGAAATTCAGACACACACGCTGGCTTGCCTAATGAAGAGAGTGCTACAGCTGATTG 795  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60  
 DB 796 ACAGCAGCAGGGTAGCCGATCTGTGCGACGACGCGCAGCTGAAATGCTGGGGCAGCA 855  
 QY 61 AlaLeuGluAlaLeuAla 66  
 DB 856 GCTTGGGAAGCAGCGGCA 873

RESULT 14  
 AAQ10002  
 ID AAQ10002 standard; DNA; 2383 BP.  
 XX AC AAQ10002;  
 DT 13-MAR-1991 (first entry)  
 DE Sequence encoding Streptococcus GX7805 protein G.  
 XX Immunoglobulins; Ig; ds.  
 OS Streptococcus sp GX7805.  
 XX Key Location/Qualifiers  
 FT CDS 577..2355  
 FT /\*tag= a

PN US4977247-A.  
 XX 11-DEC-1990.  
 XX 19-MAY-1989; 89US-0354264.  
 XX 19-MAY-1989; 89US-0354264.  
 PR 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 17-FEB-1987; 87WO-US00329.  
 PR 19-JUN-1987; 87US-0063959.  
 PR 20-JUN-1988; 88US-0209236.

Score: 155.00 Matches: 37  
 Percent Similarity: 62.12% Conservative: 4  
 Best Local Similarity: 56.06% Mismatches: 25  
 Query Match: 19.95% Indels: 0  
 DB: 11 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ06017 (1-1950)

QY 1 ValaspserProIleGluProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 DB 677 GTTGATTACCAATCGAGATACCCCAATTATTCCTAATGCTGGTGAATTAATCTT 736  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
 DB 737 CTGGGAATTCAGACACAACACTGGCTTTCGCTAATGAAGAGAGCTACAGCTGATTG 796  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
 DB 797 ACAGCAGCAGCGGTAGCCGATACTGTGGCAGCAGCGCAGCTGAANAATGCTGGGGCAGCA 856  
 QY 61 AlaLeuGluAlaLeuAla 66  
 DB 857 GCTTGGGAAGCAGCGCA 874

# RESULT 11

AAQ10001  
 ID AAQ10001 standard; DNA; 1950 BP.

AC AAQ10001;

DT 13-MAR-1991 (first entry)

DE Sequene encoding Streptococcus GX7809 protein G.

KW Immunoglobulins; Ig; ds.

OS Streptococcus sp GX7809.

XX FH Key Location/Qualifiers  
 FT CDS 578..2248  
 FT /\*tag= a

XX US4977247-A.

XX PD 11-DEC-1990.

XX PF 19-MAY-1989; 89US-0354264.

XX PR 19-MAY-1989; 89US-0354264.

XX PR 14-FEB-1986; 86US-0829354.

XX PR 23-APR-1986; 86US-0854887.

XX PR 17-FEB-1987; 87WO-US00329.

XX PR 19-JUN-1987; 87US-0063959.

XX PR 20-JUN-1988; 88US-0209236.

XX (GENE-) GENEX CORP.

XX Fahnestoc SR, Lee T, Wroble MH;

XX WPI; 1991-006758/01.

XX P-PSDB; AAR10004.

XX Immobilised protein G variants - used for detection, isolation

XX and purificn. Immunoglobulin(s) and immunoglobulin fragments

XX Disclosure: Fig 8; 52pp; English.

XX Protein G gene product may be modified allowing the variant to be

XX immobilised and exhibit different binding profiles. The bound

XX protein is useful in purification and detection of Igs and fragments.

XX Sequence 1950 BP; 706 A; 323 C; 398 G; 523 T; 0 other;

Alignment Scores:  
 Pred. No.: 1.57e-12 Length: 1950  
 Score: 155.00 Matches: 37  
 Percent Similarity: 62.12% Conservative: 4  
 Best Local Similarity: 56.06% Mismatches: 25  
 Query Match: 19.95% Indels: 0  
 DB: 12 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ10001 (1-1950)

QY 1 ValaspserProIleGluProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 DB 677 GTTGATTACCAATCGAGATACCCCAATTATTCCTAATGCTGGTGAATTAATCTT 736  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
 DB 737 CTGGGAATTCAGACACAACACTGGCTTTCGCTAATGAAGAGAGCTACAGCTGATTG 796  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
 DB 797 ACAGCAGCAGCGGTAGCCGATACTGTGGCAGCAGCGCAGCTGAANAATGCTGGGGCAGCA 856  
 QY 61 AlaLeuGluAlaLeuAla 66  
 DB 857 GCTTGGGAAGCAGCGCA 874

# RESULT 12

AAQ64644  
 ID AAQ64644 standard; DNA; 1950 BP.

AC AAQ64644;

DT 06-JAN-1995 (first entry)

DE Streptococcus Protein G gene derived from strain GX7809.

KW Streptococcus Protein G; variant; IgG binding activity;

KW immunoglobulin; Lancefield Group G; bacterial Fc receptor; ss.

OS Streptococcus sp. GX7809.

XX FH Key Location/Qualifiers  
 FT -35\_signal 465..470  
 FT /\*tag= a  
 FT -10\_signal 487..492  
 FT /\*tag= b  
 FT RBS 566..572  
 FT /\*tag= c  
 FT CDS 578..1924  
 FT /\*tag= d  
 FT /product= Protein\_G

XX US5312901-A.

XX PD 17-MAY-1994.

XX PF 14-FEB-1986; 86US-0829354.

XX PR 14-FEB-1986; 86US-0829354.

XX PR 23-APR-1986; 86US-0854887.

XX PR 19-JUN-1987; 87US-0063959.

XX PR 20-JUN-1988; 88US-0209236.

XX PR 19-JUN-1990; 90US-0540169.

XX PR 21-APR-1992; 92US-0871539.

XX (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.

XX Fahnestock SR;

XX WPI; 1994-159179/19.

XX P-PSDB; AAR53290.

XX New recombinant streptococcal protein G variants - useful for

```

Db 677 GTTGATTACCAATCGAAGATACCCCAATTATTGCTGAATGGTGGTGAATTAACATACTTT 736
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 737 CTGGGGAATTACAGACACACTGGCTTTCGTAATGAAGAGAGTCTACAGCTGATTGG 796
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 ACAGCAGCAGCGGTAGCCCATACTGTGGCAGCAGCGCAGCTGAATGCTGGGGCAGCA 856
Qy 61 AlaLeuGluAlaLeuAla 66
    ||||| ||||| |||||
Db 857 GCTTGGGAAGCAGCGGCA 874

RESULT 9
AAN91094
ID AAN91094 standard; DNA; 1950 BP.
AC AAN91094;
XX
DT 04-JUL-1990 (first entry)
XX
DE Protein G gene.
XX
KW Protein G; immunoglobulin; Fc receptor; ds.
XX
OS Streptococcus sp.
XX
FH Key Location/Qualifiers
FT -35_signal 465..470
FT /*tag= c
FT -10_signal 487..492
FT /*tag= b
FT RBS 565..571
FT /*tag= d
FT CDS 578..1918
FT /*tag= a
XX
PN W08810306-A.
XX
XX 29-DEC-1988.
PD
PF 20-JUN-1988; 88WO-US02084.
PR 19-JUN-1987; 87US-0063959.
XX
PA (GENE-) GENEX CORP.
XX
PI Fahnstock SR;
XX
DR WPI: 1989-023848/03.
DR P-PSDB; AAP95030.
XX
PT Cloned protein G variant genes -
PT expressing proteins having immunoglobulin-binding properties of
PT protein G and derived from Streptococcus sp.
XX
PS Disclosure; ; 116pp; English.
XX
CC Gene encodes protein G of non-pathogenic streptococcus sp. allowing
CC isolation of the protein and variants, useful as bacterial Fc receptors
CC eg in purification and detection of Abs., screening of hybridoma clones
CC and treatment of disease.
XX
SQ Sequence 1950 BP; 706 A; 324 C; 399 G; 521 T; 0 other;

Alignment Scores:
Pred. No.: 1.57e-12 Length: 1950
Score: 155.00 Matches: 37
Percent Similarity: 62.1% Conservative: 4
Best Local Similarity: 56.0% Mismatches: 25
Query Match: 19.95% Indels: 0
DB: 10 Gaps: 0

```

```

US-09-847-539A-6 (1-159) x AAN91094 (1-1950)
Qy 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 677 GTTGATTACCAATCGAAGATACCCCAATTATTGCTGAATGGTGGTGAATTAACATACTTT 736
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 737 CTGGGGAATTACAGACACACTGGCTTTCGTAATGAAGAGAGTCTACAGCGGATTGG 796
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 ACAGCAGCAGCGGTAGCCCATACTGTGGCAGCAGCGCAGCTGAATGCTGGGGCAGCA 856
Qy 61 AlaLeuGluAlaLeuAla 66
    ||||| ||||| |||||
Db 857 GCTTGGGAAGCAGCGGCA 874

RESULT 10
AAQ06017
ID AAQ06017 standard; DNA; 1950 BP.
XX
AC AAQ06017;
XX
DT 17-JAN-1991 (first entry)
XX
DE Sequence encoding Protein G.
XX
KW Immunoglobulin; ds.
XX
OS Streptococcus sp. Lancefield Group G strain.
XX
FH Key Location/Qualifiers
FT CDS 578..1921
FT /*tag= a
XX
PN US4956296-A.
XX
XX 11-SEP-1990.
PD
PF 20-JUN-1988; 88US-0209236.
PR 20-JUN-1988; 88US-0209236.
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87WO-US00329.
PR 19-JUN-1987; 87US-0063959.
XX
PA (GENE-) GENEX CORP.
XX
PI Fahnstock SR;
XX
DR WPI: 1990-297491/39.
DR P-PSDB; AAR07012.
XX
PT Recombinant Protein G variants - obtd. using a cloned gene
PT encoding Protein G from Streptococcus sp.; used for binding
PT immunoglobulin.
XX
PS Disclosure; Fig 3a-c; 48pp; English.
XX
CC Fragments and variants of the sequence are claimed esp. where
CC incorporated into a non-pathogenic host eg. E.coli, and expressed
CC at high levels.
CC The variants have a higher binding efficiency and capacity for
CC immunoglobulin, and may be used for purifying, detecting and
CC isolating antibodies.
XX
SQ Sequence 1950 BP; 706 A; 323 C; 398 G; 523 T; 0 other;

Alignment Scores:
Pred. No.: 1.57e-12 Length: 1950

```



ID	AAQ06018	standard: DNA; 1949 BP.
XX	AAQ06018;	
XX	17-JAN-1991	(first entry)
DT	Sequence encoding	Protein G with two active sites.
XX	Immunoglobulin; ds.	
XX	Streptococcus sp.	Lancefield Group G strain.
OS	Key	Location/Qualifiers
XX	CDS	577..1920
XX	-35_signal	/*tag= a
XX	FT	465..470
XX	-10_signal	/*tag= b
XX	FT	487..492
XX	FT	/*tag= c
XX	RBS	565..571
XX	FT	/*tag= d
XX	US4956296-A.	
XX	11-SEP-1990.	
XX	20-JUN-1988;	88US-Q209236.
XX	20-JUN-1988;	88US-Q209236.
XX	14-FEB-1986;	86US-0829354.
XX	23-APR-1986;	86US-0854887.
XX	17-FEB-1987;	87WO-US00329.
XX	19-JUN-1987;	87US-0063959.
XX	(GENE-) GENEX CORP.	
PA	Fahnestock SR;	
PI	WPI; 1990-297491/39.	
DR	P-PSDB; AAR07013.	
DR	Recombinant Protein G variants - obtd.	using a cloned gene
PT	encoding Protein G from Streptococcus sp.,	used for binding
PT	Immunoglobulin.	
XX	Disclosure; Fig 8a-c; 48pp; English.	
XX	Fragments and variants of the sequence	are claimed esp. where
CC	incorporated into a non-pathogenic host eg.	E.coli, and expressed
CC	at high levels.	
CC	The variants have a higher binding efficiency	and capacity for
CC	immunoglobulin, and may be used for purifying,	detecting and
CC	isolating antibodies.	
XX	Sequence 1949 BP; 705 A; 323 C; 398 G;	523 T; 0 other:
SQ	Alignment Scores:	
	Pred. No.:	1-57e-12
	Score:	135.00
	Percent Similarity:	62.12%
	Best Local Similarity:	56.06%
	Query Match:	19.95%
	DB:	11
	Gaps:	0
	Length:	1949
	Matches:	37
	Conservative:	4
	Mismatches:	25
	Indels:	0
	Indels:	0
	Gaps:	0
	US-09-847-539A-6 (1-159) x AAQ06018 (1-1949)	
Qy	1 ValAspSerProIleGluGlnProArgIleIleProAsnGlycylThrLeuThrAsnLeu	20
Db	676 GTTGATTCCAAATCGAAGATACCCCAATTATTCGTAATGGTGGTGAATTAATCTT	735
Qy	21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluLysArgAlaIleAspGluLeu	40
Db	736 CTGGGAATTCAGACAAACACTGGCTTTGCGTAATGAAGAGAGTCTACAGCTGATTTG	795

CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a DNA encoding partial GRAB protein from S. pyogenes strain AP49.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.  
XX  
SQ Sequence 853 BP; 295 A; 171 C; 197 G; 190 T; 0 other;

Alignment Scores:

Pred. No.: 5,8e-69 Length: 853  
Score: 536.00 Matches: 113  
Percent Similarity: 94.28% Conservativeness: 2  
Best Local Similarity: 92.62% Mismatches: 7  
Query Match: 68.98% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6 (1-159) x AAD00563 (1-853)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
DB 46 GTTGACTCACTATCGACAGCTCGAATTATTCCTCAATGGGGAACCTTAATCTT 105  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
DB 106 CTTGGCAATGCTCCAGAAACTGGCATTTACGTAATGAAGAAGAGCCATTGATGAATTA 165  
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
DB 166 AAAAAACAAGCTATTGAGGATAAAGAAGCTACGACAGCTATAGAAGCAGCAAGTTCAGAT 225  
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80  
DB 226 GCCTTTAGAGAGCATTTACGAGTCAAGCAGCGCTTTACATCAAGAGAGCTGCAGTAGTT 285  
QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
DB 286 CAATCAGATAACGCTCTAGTACGCTTACGAGCCTTACGAGCCTTACGAGCAGCGCTTTA 345  
QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120  
DB 346 CAATCAGAGAAGCTCCAGTAGTTTCAATCAGATAAGCTGCTGGTGACGCCCTTAGAAGCA 405  
QY 121 AlaAla 122  
DB 406 TTGGCG 411

RESULT 6  
AAD00561

ID AAD00561 standard; DNA; 777 BP.

XX AC AAD00561;

XX DT 29-AUG-2000 (first entry)

XX DE Streptococcus pyogenes strain KTL9 partial GRAB protein encoding DNA.

KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection; ds.

XX OS Streptococcus pyogenes.

XX FH Key Location/Qualifiers  
XX CDS 1..777

FT /tag= a  
FT /product= "GRAB protein"  
FT /note= "Does not include stop codon"  
FT /partial

XX PN W0200026240-A2.

XX PD

11-MAY-2000.

XX PF 02-NOV-1999; 99WO-GB03631.

XX PR 02-NOV-1998; 98GB-0023975.

XX PA (ACTI-) ACTINOVA LTD.

XX PI Bjorck LH, Rasmussen M;

XX WP: 2000-365572/31.

XX P-PSDB; AAY71043.

XX New alpha2M binding protein for generating a protective immune response  
XX to group A streptococcus and purifying the binding protein -

XX Claim 13; Page 63-64; 67pp; English.

XX The patent discloses a new family of proteins termed GRAB (protein G  
XX related alpha2M binding protein) from Streptococcus pyogenes which have  
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
XX protein G of group G Streptococcus. GRAB protein and peptides derived  
XX from it are used in vaccine compositions for generating a protective  
XX immune response against group A Streptococcus. Antibodies against GRAB  
XX are useful for treating Streptococcus pyogenes infections. The protein  
XX is also useful for purifying alpha2M from a sample. The present sequence  
XX is a DNA encoding partial GRAB protein from S. pyogenes strain KTL9.  
XX The protein has alpha2M binding region and is useful in vaccine  
XX composition.

SQ Sequence 777 BP; 269 A; 151 C; 178 G; 179 T; 0 other;

Alignment Scores:

Pred. No.: 1.01e-68 Length: 777  
Score: 534.00 Matches: 112  
Percent Similarity: 95.08% Conservativeness: 4  
Best Local Similarity: 91.80% Mismatches: 6  
Query Match: 68.73% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6 (1-159) x AAD00561 (1-777)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
DB 61 GTTGACTCACTATCGACAGCTCGAATTATTCCTCAATGGGGAACCTTAATCTT 120

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
DB 121 CTTGGCAATGCTCCAGAAACTGGCATTTACGTAATGAAGAAGAGCCATTGATGAATTA 180

QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
DB 181 AAAAAACAAGCTATTGAGGATAAAGAAGCTACGACAGCTATAGAAGCAGCAAGTTCAGAT 240

QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80  
DB 241 GCCTTTAGAGAGCATTTACGAGTCAAGCAGCGCTTTACATCAAGAGAGCTGCAGTAGTT 300

QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
DB 301 CAATCAGATAACGCTCTAGTACGCTTACGAGCCTTACGAGCCTTACGAGCAGCGCTTTA 360

QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120  
DB 361 CAATCAGAGAAGCTCGCGTTGTTAAAGCGGATAACGCTGCTAGTCACACTTTTAGAAGCA 420

QY 121 AlaAla 122

DB 421 TTGGCG 426

RESULT 7  
AAQ06018

Db 454 AAAGAGAAAGACAAACGTTAATACCTTCTTACAACTGGTGAAGAG 501  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 4

AAD00562

ID AAD00562 standard; DNA; 469 BP.

XX AC AAD00562;

XX 29-AUG-2000 (first entry)

XX Streptococcus pyogenes strain AP1 partial GRAB protein encoding DNA.

XX GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

KW immune response; Streptococcus pyogenes infection; ds.

XX Streptococcus pyogenes.

OS Streptococcus pyogenes.

XX Key Location/Qualifiers

FT 1..468

FT CDS

FT /\*tag= a

FT /product= "GRAB protein"

FT /partial

XX WO200026240-A2.

XX 11-MAY-2000.

XX 02-NOV-1999; 99WO-GB03631.

XX 02-NOV-1998; 98GB-0023975.

XX (ACTI-) ACTINOVA LTD.

XX Bjorck LH, Rasmussen M;

XX WPI; 2000-365572/31.

DR P-PSDB; AAY71044.

XX New alpha2M binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein

XX Claim 13; Page 64; 67pp; English.

XX The patent discloses a new family of proteins termed GRAB (protein G

CC related alpha2M binding protein) from Streptococcus pyogenes which have

CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to

CC protein G of group A Streptococcus. GRAB protein and peptides derived

CC from it are used in vaccine compositions for generating a protective

CC immune response against group A Streptococcus. Antibodies against GRAB

CC are useful for treating Streptococcus pyogenes infections. The protein

CC is also useful for purifying alpha2M from a sample. The present sequence

CC is a DNA encoding partial GRAB protein from S. pyogenes strain API.

CC The protein has alpha2M binding region and is useful in vaccine

CC composition.

XX SQ Sequence 469 BP; 180 A; 90 C; 99 G; 100 T; 0 other;

Alignment Scores:

Pred. No.: 9.74e-101 Length: 469

Score: 750.00 Matches: 154

Percent Similarity: 99.35% Conservative: 0

Best Local Similarity: 99.35% Mismatches: 1

Query Match: 96.53% Indels: 0

DB: 21 Gaps: 0

US-09-847-539A-6 (1-159) x AAD00562 (1-469)

Qy 2 AspSerProIleGluInProArgIleProAsnGlyGlyThrLeuThrAsnLeuLeu 21

Db 1 GACTCACCCTATCGACAGCCCTAGATTATTCCTCAATGGCGGACCTTANTTAATCTCTT 60

QY 22 GlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys 41  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 GGCAATGCTCCAGAAAACTGGCATTTACGTAATCAAGAAAGAGCCATGTGATGATTTAAA 120  
 QY 42 LysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAla 61  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 AAACAAGCTATTGAGGATAAGGAAGCTACGACAGCTATAGAAGCAGCAAGTTCAGATGCC 180  
 QY 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 81  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 181 TTAGAACCATTTAGCGGATCAACAGACGCTTTACAATCAGAAGAAGCTGCGGTTGTTAAA 240  
 QY 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 CGCGATAACGCTGCTAGTAGCGCTTTAGAAGCAATTTGGCGATCAACAGAGCGCTTTACAA 300  
 QY 102 SerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTtpGluLysAla 121  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 301 TCAGAAGAAGCTGAAGTAGTTCAATCAGATAACGCTGCTAGTGACGCTGGGAAAAAGCA 360  
 QY 122 AlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLysLys 141  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 361 GCAACTCCATCGCTTTAGATGTTAAGAAAACCTAAAGATACAAAACCTGTAGTTAAAAA 420  
 QY 142 GluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 421 GAAGAAAGACAAAACGTTAATACCTTCTTACAACTGGTGAAGAG 465

RESULT 5

AAD00563

ID AAD00563 standard; DNA; 853 BP.

XX AC AAD00563;

XX 29-AUG-2000 (first entry)

XX Streptococcus pyogenes strain AP49 partial GRAB protein encoding DNA.

XX GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

KW immune response; Streptococcus pyogenes infection; ds.

XX Streptococcus pyogenes.

OS Streptococcus pyogenes.

XX Key Location/Qualifiers

FT 37..852

FT CDS

FT /\*tag= a

FT /product= "GRAB protein"

FT /partial

XX WO200026240-A2.

XX 11-MAY-2000.

XX 02-NOV-1999; 99WO-GB03631.

XX 02-NOV-1998; 98GB-0023975.

XX (ACTI-) ACTINOVA LTD.

XX Bjorck LH, Rasmussen M;

XX WPI; 2000-365572/31.

DR P-PSDB; AAY71045.

XX New alpha2M binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein

XX Claim 13; Page 64; 67pp; English.

XX The patent discloses a new family of proteins termed GRAB (protein G

CC related alpha2M binding protein) from Streptococcus pyogenes which have

CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to

CC protein G of group A Streptococcus. GRAB protein and peptides derived

CC from it are used in vaccine compositions for generating a protective

CC immune response against group A Streptococcus. Antibodies against GRAB

CC are useful for treating Streptococcus pyogenes infections. The protein

CC is also useful for purifying alpha2M from a sample. The present sequence

CC is a DNA encoding partial GRAB protein from S. pyogenes strain API.

CC The protein has alpha2M binding region and is useful in vaccine

CC composition.

XX SQ Sequence 469 BP; 180 A; 90 C; 99 G; 100 T; 0 other;

Alignment Scores:

Pred. No.: 9.74e-101 Length: 469

Score: 750.00 Matches: 154

Percent Similarity: 99.35% Conservative: 0

Best Local Similarity: 99.35% Mismatches: 1

Query Match: 96.53% Indels: 0

DB: 21 Gaps: 0

US-09-847-539A-6 (1-159) x AAD00562 (1-469)

CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2M from a sample. The present sequence  
 CC is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.  
 XX  
 SQ Sequence 764 BP; 279 A; 131 C; 159 G; 195 T; 0 other;

## Alignment Scores:

Pred. No.: 2.03e-104 Length: 764  
 Score: 777.00 Matches: 159  
 Percent Similarity: 100.00% Conservativeness: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-847-539A-6 (1-159) x AAD00559 (1-764)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 DB 159 GTTGACTCACCTATCGAAGCCTCGAATATTTCCAAATGGCGGAACCTTAATCTT 218  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
 DB 219 CTTGGCATGCTCCAGAAAACCTGGCATTTACGTATATGAAGAAGAGCCATTGATGATTA 278  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
 DB 279 AAAAACAAGCTATTGAGATAAAGAAGCTACGACAGCTATAGAAGCAGCAAGTTTCAGAT 338  
 QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValVal 80  
 DB 339 GCCTTAGAAGCATTAACGGATCAACAGACGCTTTACAATCAGAAGAGCTGCGGTGTT 398  
 QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
 DB 399 AAGCGGATAACGCTGCTAGTCAGCGCTTTAGAAGCATTTGGCGGATCAACAGACGCTTTA 458  
 QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120  
 DB 459 CAATCAGAAGAAGCTGAAGTAGTTCAATCAGATAACGCTGCTAGTGACGCTGGGAAAAA 518  
 QY 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140  
 DB 519 GCAGCACTCCCACTGCTTAGATGTTAAGAAACATAAAGATACAAACCTGTAGTTAA 578  
 QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsnPro 159  
 DB 579 AAGAAGAAGACAAACAGCTTAATACCTTCCTACAACTGGTGAGAGCTTAACCCA 635

## RESULT 3

AAD00564  
 ID AAD00564 standard; DNA; 504 BP.  
 XX  
 AC AAD00564;

XX 29-AUG-2000 (first entry)

DE Streptococcus pyogenes strain KTL3 partial GRAB protein encoding DNA.  
 KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
 KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
 KW Immune response; Streptococcus pyogenes infection; ds.

OS Streptococcus pyogenes.

XX Key Location/Qualifiers  
 FH CDS  
 FT 1..504  
 FT /\*tag= a  
 FT /product= \*GRAB protein\*  
 FT /partial

PN WO200026240-A2.  
 XX 11-MAY-2000.  
 PD 02-NOV-1999; 99WO-GB03631.  
 PF 02-NOV-1998; 98GB-0023975.  
 PR (ACTI-) ACTINOVA LTD.  
 XX Bjorck LH, Rasmussen M;  
 XX WPI: 2000-365572/31.  
 DR P-PSDB; AAY71046.  
 XX New alpha2M binding protein for generating a protective immune response  
 PT to group A streptococcus and purifying the binding protein -  
 PS Claim 13; Page 65; 67pp; English.  
 CC The patent discloses a new family of proteins termed GRAB (protein G  
 CC related alpha2M binding protein) from Streptococcus pyogenes which have  
 CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
 CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2M from a sample. The present sequence  
 CC is a DNA encoding partial GRAB protein from S. pyogenes strain KTL3.  
 CC The protein has alpha2M binding region and is useful in vaccine  
 CC composition.  
 XX Sequence 504 BP; 188 A; 97 C; 108 G; 111 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 3.65e-102 Length: 504  
 Score: 760.00 Matches: 156  
 Percent Similarity: 100.00% Conservativeness: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 97.81% Indels: 0  
 DB: 21 Gaps: 0  
 US-09-847-539A-6 (1-159) x AAD00564 (1-504)  
 QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 DB 34 GTTGACTCACCTATCGAAGCCTCGAATATTTCCAAATGGCGGAACCTTAATCTT 93  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
 DB 94 CTTGGCATGCTCCAGAAAACCTGGCATTTACGTATATGAAGAAGAGCCATTGATGATTA 153  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
 DB 154 AAAAAACAAGCTATTGAGATAAAGAAGCTACGACAGCTATAGAAGCAGCAAGTTTCAGAT 213  
 QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80  
 DB 214 GCCTTAGAAGCATTAACGGATCAACAGACGCTTTACAATCAGAAGAAGCTGCGGTGTT 273  
 QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
 DB 274 AAGCGGATAACGCTGCTAGTGACGCTTAGAAGCATTTGGCGGATCAACAGACGCTTTA 333  
 QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120  
 DB 334 CAATCAGAAGAAGCTGAAGTAGTTCAATCAGATAACGCTGCTAGTGACGCTGGGAAAAA 393  
 QY 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140  
 DB 394 GCAGCAACTCCCACTGCTTAGATGTTAAGAAACATAAAGATACAAACCTGTAGTTAA 453  
 QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156

```
FT      /*tag= a
FT      /product= "GRAB protein"
FT      1..99
FT      sig_peptide
FT      /*tag= b
FT      100..651
FT      mat_peptide
FT      /*tag= c
FT      /product= "Mature GRAB protein"
XX
XX WO200026240-A2.
XX
XX PD 11-MAY-2000.
XX
XX PF 02-NOV-1999; 99WO-GB03631.
XX
XX PR 02-NOV-1998; 98GB-0023975.
XX
XX PA (ACTI-) ACTINOVA LTD.
XX
XX PI Bjorck LH, Rasmussen M;
XX
XX KW WPI; 2000-365572/31.
XX
XX DR P-PSDB; AAY71042.
XX
XX PT New alpha2M binding protein for generating a protective immune response
XX      to group A streptococcus and purifying the binding protein -
XX
XX PS Claim 13; Page 63; 67pp; English.
XX
XX CC The patent discloses a new family of proteins termed GRAB (protein G
XX      related alpha2M binding protein) from Streptococcus pyogenes which have
XX      the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX      protein G of group G Streptococcus. GRAB protein and peptides derived
XX      from it are used in vaccine compositions for generating a protective
XX      immune response against group A Streptococcus. Antibodies against GRAB
XX      are useful for treating Streptococcus pyogenes infections. The protein
XX      is also useful for purifying alpha2M from a sample. The present sequence
XX      is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.
XX
XX SQ Sequence 654 BP; 234 A; 118 C; 143 G; 159 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 1.65e-104 Length: 654
XX Score: 777.00 Matches: 159
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-847-539A-6 (1-159) x AAD00560 (1-654)
XX
QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
DB 100 GTTGACTCACCTATCGACAGCTCGAATTATTCCTCAATGGCGGAACCTTAACCTACTT 159
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
DB 160 CTTGCAATGTCTCCAGAAAACCTGCGATTACGTAATGAAGAAGAGCCATTGATGAATTA 219
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
DB 220 AAAAACAAGCTTATTGAGCATAAAGAGCTAGCAGCATATAGAAGCAGCAGATTCACAT 279
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
DB 280 GCCTTAGAGCATTACGGATCAACACAGCGCTTACAAATCAGAAGAAGCTGCGGTGTT 339
QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
DB 340 AAAGCGGATACGCTGCTAGTCAGCGCTTAGAAGCATTTGGCGGATCAACACAGCGTTTA 399
QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
DB 400 CAATCAGAAGAAGCTGAAGTAGTTCAATCAGATAACGCTGCTAGTAGTGACGCTGGGAAAA 459
```

```
QY 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140
DB 460 GCAGCAACTCCCAATCGCTTTAGATGTTAAGAAACTAAAGATACAAAACCTGTAGTTAAA 519
QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsnPro 159
DB 520 AAAGAAGAAGACAAACGTTTAATACCTTCTCTACAACTGGTGAAGAGTCTAACCCA 576
XX
XX RESULT 2
XX AAD00559
XX ID AAD00559 standard; DNA; 764 BP.
XX
XX AC AAD00559;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX DE Streptococcus pyogenes strain SF370 GRAB protein encoding DNA.
XX
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;
XX      alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX      immune response; Streptococcus pyogenes infection; ds.
XX
XX OS Streptococcus pyogenes.
XX
XX FH Key Location/Qualifiers
XX CDS 60..713
XX FT /*tag= a
XX FT /product= "GRAB protein"
XX FT sig_peptide 60..158
XX FT /*tag= b
XX FT mat_peptide 159..710
XX FT /*tag= c
XX FT /product= "Mature GRAB protein"
XX FT primer_bind complement (101..124)
XX FT /*tag= d
XX FT primer_bind complement (101..127)
XX FT /*tag= e
XX FT primer_bind complement (160..184)
XX FT /*tag= f
XX FT primer_bind complement (563..594)
XX FT /*tag= g
XX FT primer_bind complement (605..626)
XX FT /*tag= h
XX FT primer_bind complement (605..626)
XX
XX PN WO200026240-A2.
XX
XX PD 11-MAY-2000.
XX
XX PF 02-NOV-1999; 99WO-GB03631.
XX
XX PR 02-NOV-1998; 98GB-0023975.
XX
XX PA (ACTI-) ACTINOVA LTD.
XX
XX PI Bjorck LH, Rasmussen M;
XX
XX KW WPI; 2000-365572/31.
XX
XX DR P-PSDB; AAY71042.
XX
XX PT New alpha2M binding protein for generating a protective immune response
XX      to group A streptococcus and purifying the binding protein -
XX
XX PS Example 1; Fig 2B; 67pp; English.
XX
XX CC The patent discloses a new family of proteins termed GRAB (protein G
XX      related alpha2M binding protein) from Streptococcus pyogenes which have
XX      the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:10:56 : Search time 229.572 Seconds  
(without alignments)  
1189.122 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQPRIIPNGTGLTNL.....KKEERQNVNLTPTTGEESNP 159

Scoring table: BLOSUM62

Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_092548\_3413/app\_query.fasta\_1.526  
-DB=N\_Geneseq\_032802 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsm62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=pto -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847533@cgn\_1.1.48 -runat\_10102002\_092548\_3413 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAL -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FPGAPOP=6 -FPGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	777	100.0	654	21	Streptococcus pyog
2	777	100.0	764	21	Streptococcus pyog
3	760	97.8	504	21	Streptococcus pyog
4	750	96.5	469	21	Streptococcus pyog
5	536	69.0	853	21	Streptococcus pyog
6	534	68.7	777	21	Streptococcus pyog
7	155	19.9	1949	11	Sequence encoding
8	155	19.9	1950	10	Protein G gene. S
9	155	19.9	1950	10	Protein G gene. S
10	155	19.9	1950	11	Sequence encoding
11	155	19.9	1950	12	Sequence encoding
12	155	19.9	1950	15	Streptococcus prot
13	155	19.9	2383	11	Sequence encoding
14	155	19.9	2383	12	Sequence encoding
15	155	19.9	2383	15	Streptococcus prot
16	142	18.3	1576	21	Streptococcus stra
17	105	13.5	798	10	Protein G gene var
18	105	13.5	798	11	Sequence encoding
19	105	13.5	798	12	Type 4 GX7809 prot
20	105	13.5	798	15	IGC-binding Strept
21	93	12.0	1380	23	DNA encoding novel
22	93	12.0	1383	23	DNA encoding novel
23	91	11.7	1555	16	S. dysgalactiae ma
24	91	11.7	1555	16	S. dysgalactiae ma
25	90	11.6	1469	8	Sequence encoding
26	90	11.6	2121	23	Drosophila melanog
27	90	11.6	1145	23	Drosophila melanog
28	86	11.1	3411	23	DNA encoding novel
29	85	10.9	2526	16	S. dysgalactiae mi
30	84	10.8	2129	23	Drosophila melanog
31	84	10.8	2129	23	Drosophila melanog
32	83	10.7	5366	23	ABLO3677
33	83	10.7	603	16	Expression cassett
34	83	10.7	627	16	Chimeric synthe
35	83	10.7	812	21	Synthetic protein
36	83	10.7	2561	23	Arabidopsis thalia
37	81	10.4	2010	24	DNA encoding novel
38	81	10.4	3069	23	Streptococcus dysg
39	81	10.4	3069	23	DNA encoding novel
40	77	9.9	782	23	DNA encoding novel
41	77	9.9	1779	23	DNA encoding novel
42	76	9.8	441	22	S. epidermidis ope
43	76	9.8	3600	22	S. epidermidis gen
44	76	9.8	10136	17	Kinetochore protei
45	75	9.7	1290	22	CPE 30 coding sequ

ALIGNMENTS

RESULT 1  
AAD00560  
ID AAD00560 standard; DNA; 654 BP.  
XX  
AC AAD00560;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 GRAB protein coding region.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW Immune response; Streptococcus pyogenes infection; ds.  
XX  
OS Streptococcus pyogenes.  
XX  
FH Key  
FT CDS Location/Qualifiers  
1..654

**THIS PAGE BLANK (uspto)**

; LIBRARY: LAMBDA GEM(TM)-11 LIBRARY OF KENNETH NARVA  
; CLONE: PS69D1A  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..395  
US-08-049-783-6

Query Match 31.8%; Score 41; DB 1; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.9e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAIVVKADN 26  
| | : | | : | : | : |  
Db 231 EELLKVVDDLKKELEAAIKAEK 252

Search completed: October 13, 2002, 04:50:07  
Job time : 5.34225 secs



COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01676A  
FILING DATE: 19930226  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 07/842,644  
FILING DATE: 02-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooper, Iver P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: STAMSBROOK 1  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 370 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-01676A-4

Query Match 31.8%; Score 41; DB 5; Length 370;  
Best Local Similarity 43.5%; Pred. No. 1.7e+02;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 SDALADADTDALQSEEAAYVK 23  
DB 343 ADSLMQARQVSRLESQQAALPK 365

RESULT 39  
US-07-876-280-9  
Sequence 9, Application US/07876280  
Patent No. 5262158  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel M.  
APPLICANT: Cannon, Raymond J.C.  
APPLICANT: Bagley, Angela L.  
TITLE OF INVENTION: No. 5262158el Bacillus thuringiensis Isolates for  
CONTROLLING ACARIDES  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/876,280  
FILING DATE: 19920430  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/S 104  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 395 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BACILLUS THURINGIENSIS  
INDIVIDUAL ISOLATE: PS69D1  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2317) NRRL B-18816  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..395  
US-07-876-280-9

Query Match 31.8%; Score 41; DB 1; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.9e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADOTDALQSEEAAYVKADN 26  
DB 231 EELLKKVDDLKKELEAAIKAE 252

RESULT 40  
US-08-049-783-6  
Sequence 6, Application US/08049783  
Patent No. 5439881  
GENERAL INFORMATION:  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Schwab, George E.  
APPLICANT: Payne, Jewel M.  
TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active  
TOXINS CLONED FROM BACILLUS THURINGIENSIS ISOLATES  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jeff Lloyd  
STREET: 2421 N.W. 41st Street  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/049,783  
FILING DATE: 19930419  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lloyd, Jeff  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6 (PS69D1):  
SEQUENCE CHARACTERISTICS:  
LENGTH: 395 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BACILLUS THURINGIENSIS  
INDIVIDUAL ISOLATE: PS69D1  
IMMEDIATE SOURCE:





QY 2 DALEADOTDQALQSEEAAYVKAD 25  
||||| ||| : |||  
Db 1356 DALEAL---DRALDRETTVVAD 1376

RESULT 27  
US-08-710-749-8  
; Sequence 8, Application US/08710749  
; Patent No. 5955089  
; GENERAL INFORMATION:  
; APPLICANT: Briles, David E.  
; APPLICANT: Hollingshead, Susan  
; APPLICANT: Becker, Robert  
; TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patulin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/710,749  
; FILING DATE: 20-SEP-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frommer, William S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454312-2074  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 102 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: n/a  
; TOPOLOGY: linear  
; MOLECULE TYPE: amino acid  
US-08-710-749-8

Query Match 32.6%; Score 42; DB 2; Length 102;  
Best Local Similarity 45.5%; Pred. No. 27;  
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 4 LEALADOTDQALQSEEAAYVKAD 25  
||| : ||| : |||  
Db 36 LEELSKIDELDAEIAKLNKDD 57

RESULT 28  
US-09-521-668B-18  
; Sequence 18, Application US/09521668B  
; Patent No. 6303383  
; GENERAL INFORMATION:  
; APPLICANT: NAKAMURA, JUN  
; APPLICANT: KANNO, SOHEI  
; APPLICANT: KIMURA, EIICHIRO  
; APPLICANT: MATSUI, KAUIKO  
; APPLICANT: NAKAMATSU, TSUYOSHI  
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA  
; FILE REFERENCE: 0010-1093-0  
; CURRENT APPLICATION NUMBER: US/09/521,668B  
; CURRENT FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: JP 11-69896  
; PRIOR FILING DATE: 1999-03-16

; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 18  
; LENGTH: 427  
; TYPE: PRT  
; ORGANISM: Brevibacterium lactofermentum  
US-09-521-668B-18

Query Match 32.6%; Score 42; DB 4; Length 427;  
Best Local Similarity 47.4%; Pred. No. 1.5e+02;  
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 LEALADOTDQALQSEEAAYV 22  
| : ||| : ||| : |||  
Db 256 LDMLAQSDAGEDMDAVLV 274

RESULT 29  
US-09-521-668B-20  
; Sequence 20, Application US/09521668B  
; Patent No. 6303383  
; GENERAL INFORMATION:  
; APPLICANT: NAKAMURA, JUN  
; APPLICANT: KANNO, SOHEI  
; APPLICANT: KIMURA, EIICHIRO  
; APPLICANT: MATSUI, KAUIKO  
; APPLICANT: NAKAMATSU, TSUYOSHI  
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA  
; FILE REFERENCE: 0010-1093-0  
; CURRENT APPLICATION NUMBER: US/09/521,668B  
; CURRENT FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: JP 11-69896  
; PRIOR FILING DATE: 1999-03-16  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 20  
; LENGTH: 427  
; TYPE: PRT  
; ORGANISM: Brevibacterium lactofermentum  
US-09-521-668B-20

Query Match 32.6%; Score 42; DB 4; Length 427;  
Best Local Similarity 47.4%; Pred. No. 1.5e+02;  
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 LEALADOTDQALQSEEAAYV 22  
| : ||| : ||| : |||  
Db 256 LDMLAQSDAGEDMDAVLV 274

RESULT 30  
US-09-413-814-67  
; Sequence 67, Application US/09413814  
; Patent No. 6225064  
; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Beyer, Stefan  
; APPLICANT: Bioecker, Helmut  
; APPLICANT: Brandt, Petra  
; APPLICANT: Cino, Paul M  
; APPLICANT: Dougherty, Brian A  
; APPLICANT: Goldberg, Steven L  
; APPLICANT: Hofle, Gerhard  
; APPLICANT: Mueller, Joachim  
; APPLICANT: Reichenbach, Hans  
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or  
; FILE REFERENCE: PCT/US 99/23535  
; CURRENT APPLICATION NUMBER: US/09/413,814  
; CURRENT FILING DATE: 1999-10-07  
; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
; EARLIER FILING DATE: 1998-10-09

```
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238.557
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-238-557-2

Query Match 33.3%; Score 43; DB 4; Length 418;
Best Local Similarity 33.3%; Pred. No. 1e+02;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 2 DALEALADQTDALQSEEAAYVKADNA 28
DB 167 DHOETGLDGLAEIAIPQKAGIFKAGKA 193

RESULT 25
US-08-764-100-25
; Sequence 25, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.
; TITLE OF INVENTION: Improvements in or Relating to Organic
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sandoz Agro, Inc
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764.100
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,064
; FILING DATE:
; APPLICATION NUMBER: US 08/032,235
```

```
; FILING DATE: 17-MAR-1993
; APPLICATION NUMBER: GB 9206016.9
; FILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5773700ris, Allen E.
; REGISTRATION NUMBER: 34,490
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-764-100-25

Query Match 33.3%; Score 43; DB 1; Length 744;
Best Local Similarity 38.5%; Pred. No. 2.1e+02;
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 SDALEALADQTDALQSEEAAYVKADN 26
DB 432 SERLEPRLEILEALASERVALLEASN 457

RESULT 26
US-08-804-227C-5
; Sequence 5, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804.227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1611 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-804-227C-5

Query Match 32.9%; Score 42.5; DB 2; Length 1611;
Best Local Similarity 54.2%; Pred. No. 6.1e+02;
Matches 13; Conservative 1; Mismatches 7; Indels 3; Gaps 1;
```

APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd Q  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: GM10088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 407 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-934-846-4

Query Match 33.3%; Score 43; DB 2; Length 407;  
Best Local Similarity 33.3%; Pred. No. 1e+02;  
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 2 DALEALADQTDALQSEEAHVVKADNAA 28  
| | | | | : | : | : | | |  
DB 167 DHOETLGDSLEATAEQKAGIFKAGKKA 193

## RESULT 22

US-09-238-557-4  
Sequence 4, Application US/09238557  
Patent No. 6165472  
GENERAL INFORMATION:

APPLICANT: Pearson, Stewart C.  
APPLICANT: Greenwood, Rebecca C.  
TITLE OF INVENTION: NOVEL FOIC  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/238,557  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/934,846  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd Q  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: GM10088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222  
TELEX:

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 407 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-238-557-4

Query Match 33.3%; Score 43; DB 4; Length 407;  
Best Local Similarity 33.3%; Pred. No. 1e+02;

QY 2 DALEALADQTDALQSEEAHVVKADNAA 28  
| | | | | : | : | : | | |  
DB 167 DHOETLGDSLEATAEQKAGIFKAGKKA 193

## RESULT 23

US-08-934-846-2  
Sequence 2, Application US/08934846  
Patent No. 5882898  
GENERAL INFORMATION:  
APPLICANT: Pearson, Stewart C.  
APPLICANT: Greenwood, Rebecca C.  
TITLE OF INVENTION: NOVEL FOIC  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/934,846  
FILING DATE:

CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd Q  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: GM10088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222  
TELEX:

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 418 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-934-846-2

Query Match 33.3%; Score 43; DB 2; Length 418;  
Best Local Similarity 33.3%; Pred. No. 1e+02;

Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 2 DALEALADQTDALQSEEAHVVKADNAA 28  
| | | | | : | : | : | | |  
DB 167 DHOETLGDSLEATAEQKAGIFKAGKKA 193

## RESULT 24

US-09-238-557-2  
Sequence 2, Application US/09238557  
Patent No. 6165472  
GENERAL INFORMATION:

APPLICANT: Pearson, Stewart C.  
APPLICANT: Greenwood, Rebecca C.  
TITLE OF INVENTION: NOVEL FOIC  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia

```
RESULT 19
US-08-235-838-7
; Sequence 7, Application US/08235838
; Patent No. 5571894
; GENERAL INFORMATION:
; APPLICANT: Wels, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5571894man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,838
; FILING DATE: TBA
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-235-838-7

Query Match 33.7%; Score 43.5; DB 1; Length 711;
Best Local Similarity 31.7%; Pred. No. 1.6e+02;
Matches 13; Conservative 6; Mismatches 9; Indels 13; Gaps 1;

QY 1 SDALEALADQTD-----LQSEEAAVKADNAA 28
;::: || || || | | | | | | | |
Db 555 NDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAA 595

RESULT 20
US-08-465-473B-7
; Sequence 7, Application US/08465473B
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Wels, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5939531man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
```

```
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NOVARTIS Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901-6940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,473B
; FILING DATE: 5 June 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pfeiffer, Hesna J.
; REGISTRATION NUMBER: 22,640
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)522 6940
; TELEFAX: (908)522 6955
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-473B-7

Query Match 33.7%; Score 43.5; DB 2; Length 711;
Best Local Similarity 31.7%; Pred. No. 1.6e+02;
Matches 13; Conservative 6; Mismatches 9; Indels 13; Gaps 1;

QY 1 SDALEALADQTD-----LQSEEAAVKADNAA 28
;::: || || || | | | | | | | |
Db 555 NDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAA 595

RESULT 21
US-08-934-846-4
; Sequence 4, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL FOLC
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,846
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
```

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/100,708  
FILING DATE: July 29, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Wong, Wean Khing  
REGISTRATION NUMBER: 33,561  
REFERENCE/DOCKET NUMBER: 5324 US.P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 938-3517  
TELEFAX: (708) 938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 471 amino acid residues  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM:  
US-08-657-392-31

Query Match 33.7% Score 43.5; DB 2; Length 471;  
Best Local Similarity 31.7% Pred. No. 1e+02;  
Matches 13; Conservative 6; Mismatches 9; Indels 13; Gaps 1;

Oy 1 SDALEALADQTD-----LQSEEAHVYKADNAA 28  
Db 315 NDSVPTLAQMTDKATELLSKNEKGFLLQVAGASIDKQDHA 355

## RESULT 17

PCT-US94-02539-31  
Sequence 31, Application PC/TUS9402539  
GENERAL INFORMATION:  
APPLICANT: Brate, E.M.  
APPLICANT: Brennan, C.A.  
APPLICANT: Bridon, D.P.  
APPLICANT: Jaffe, K.D.  
APPLICANT: Krafft, G.A.  
APPLICANT: Mandeckl, W.  
APPLICANT: March, S.C.  
APPLICANT: Russell, J.R.  
APPLICANT: Yue, V.T.  
TITLE OF INVENTION: Genetically Engineered Enzymes  
TITLE OF INVENTION: And Their  
TITLE OF INVENTION: Conjugates For Diagnostic Assays  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES  
STREET: One Abbott Park Road  
CITY: Abbott Park  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: SoftPC  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/02539  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wong, Wean Khing  
REGISTRATION NUMBER: 33,561  
REFERENCE/DOCKET NUMBER: 5324.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 938-3517  
TELEFAX: (708) 938-2623  
TELEX:

INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 471 amino acid residues  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM:  
PCT-US94-02539-31

Query Match 33.7% Score 43.5; DB 5; Length 471;  
Best Local Similarity 31.7% Pred. No. 1e+02;  
Matches 13; Conservative 6; Mismatches 9; Indels 13; Gaps 1;

Oy 1 SDALEALADQTD-----LQSEEAHVYKADNAA 28  
Db 315 NDSVPTLAQMTDKATELLSKNEKGFLLQVAGASIDKQDHA 355

## RESULT 18

US-08-928-213B-8  
Sequence 8, Application US/08928213B  
Patent No. 6238905  
GENERAL INFORMATION:  
APPLICANT: McHenry, Charles S.  
Seville, Mark  
Cull, Millard G.  
TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III  
HOLOENZYME  
NUMBER OF SEQUENCES: 195  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,213B  
FILING DATE: 12-Sep-1997  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: MacKnight, Kamrin T.  
REGISTRATION NUMBER: 38,230  
REFERENCE/DOCKET NUMBER: ENZYCO-02550  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-705-8410  
TELEFAX: 415-397-8338  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 528 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-08-928-213B-8

Query Match 33.7% Score 43.5; DB 4; Length 528;  
Best Local Similarity 46.2% Pred. No. 1.2e+02;  
Matches 12; Conservative 7; Mismatches 6; Indels 1; Gaps 1;

Oy 2 DALEALADQTDALQSEEAHVYKADNA 27  
Db 314 EAMERLARRSDAL-SLEVALLEAGRA 338



```
? APPLICANT: March, S.C.
? APPLICANT: Russell, J.R.
? APPLICANT: Yue, V.T.
? TITLE OF INVENTION: Genetically Engineered Enzymes And Their
? CONJUGATES FOR DIAGNOSTIC ASSAYS
? NUMBER OF SEQUENCES: 34
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: ABBOTT LABORATORIES
? STREET: One Abbott Park Road
? CITY: Abbott Park
? STATE: Illinois
? COUNTRY: USA
? ZIP: 60064-3500
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: SoftPC
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/657,392
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/100,708
? FILING DATE: July 29, 1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Wong, Wean Khing
? REGISTRATION NUMBER: 33,561
? REFERENCE/DOCKET NUMBER: 5324.US.PI
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (708) 938-3517
? TELEFAX: (708) 938-2623
? TELEX:
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 449 amino acid residues
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: unknown
? MOLECULE TYPE: protein
? ORIGINAL SOURCE:
? ORGANISM:
? PCT-US94-02539-2
?
? Query Match 33.7%; Score 43.5; DB 2; Length 449;
? Best Local Similarity 31.7%; Pred. No. 95;
? Matches 13; Conservative 6; Mismatches 9; Indels 13; Gaps 1;
?
? QY 1 SDALEALADQTD-----LQSEEAAYVKADNAA 28
? Db 293 NDSVPTLAQWTKAIELLSKNEKGFFLQVEGASIDKODHAA 333
?
? RESULT 15
? PCT-US94-02539-2
? SEQUENCE 2, Application PC/TUS9402539
? GENERAL INFORMATION:
? APPLICANT: Brate, E.M.
? APPLICANT: Brennan, C.A.
? APPLICANT: Bridon, D.P.
? APPLICANT: Jaffe, K.D.
? APPLICANT: Krafft, G.A.
? APPLICANT: Mandeckl, W.
? APPLICANT: March, S.C.
? APPLICANT: Russell, J.R.
? APPLICANT: Yue, V.T.
? TITLE OF INVENTION: Genetically Engineered Enzymes
? AND THEIR CONJUGATES FOR DIAGNOSTIC ASSAYS
? NUMBER OF SEQUENCES: 34
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: ABBOTT LABORATORIES
? STREET: One Abbott Park Road
```

```
? CITY: Abbott Park
? STATE: Illinois
? COUNTRY: USA
? ZIP: 60064-3500
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: SoftPC
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US94/02539
? FILING DATE:
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Wong, Wean Khing
? REGISTRATION NUMBER: 33,561
? REFERENCE/DOCKET NUMBER: 5324.PC.01
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (708) 938-3517
? TELEFAX: (708) 938-2623
? TELEX:
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 449 amino acid residues
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: unknown
? MOLECULE TYPE: protein
? ORIGINAL SOURCE:
? ORGANISM:
? PCT-US94-02539-2
?
? Query Match 33.7%; Score 43.5; DB 5; Length 449;
? Best Local Similarity 31.7%; Pred. No. 95;
? Matches 13; Conservative 6; Mismatches 9; Indels 13; Gaps 1;
?
? QY 1 SDALEALADQTD-----LQSEEAAYVKADNAA 28
? Db 293 NDSVPTLAQWTKAIELLSKNEKGFFLQVEGASIDKODHAA 333
?
? RESULT 16
? US-08-657-392-31
? SEQUENCE 31, Application US/08657392
? Patent No. 5843634
? GENERAL INFORMATION:
? APPLICANT: Brate, E.M.
? APPLICANT: Brennan, C.A.
? APPLICANT: Bridon, D.P.
? APPLICANT: Jaffe, K.D.
? APPLICANT: Krafft, G.A.
? APPLICANT: Mandeckl, W.
? APPLICANT: March, S.C.
? APPLICANT: Russell, J.R.
? APPLICANT: Yue, V.T.
? TITLE OF INVENTION: Genetically Engineered Enzymes And Their
? CONJUGATES FOR DIAGNOSTIC ASSAYS
? NUMBER OF SEQUENCES: 34
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: ABBOTT LABORATORIES
? STREET: One Abbott Park Road
? CITY: Abbott Park
? STATE: Illinois
? COUNTRY: USA
? ZIP: 60064-3500
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: SoftPC
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/657,392
? FILING DATE:
```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/041173  
FILING DATE: 15-APR-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/049,282  
FILING DATE: 16-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 12176-34-1PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 382 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-04173-2

Query Match 33.7% Score 43.5; DB 5; Length 382;  
Best Local Similarity 52.4% Pred. No. 78;  
Matches 11; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

QY 4 LEALADQTDALQSEEAAYVK 23  
||| I:|:| I:|:| I:|:| I:|:|  
DB 243 LEARNQTEATLQTKDTGTVK 263

RESULT 12  
US-08-537-715-4  
Sequence 4, Application US/08537715  
Patent No. 5910627  
GENERAL INFORMATION:  
APPLICANT: Chuck, George S.  
APPLICANT: Dooner, Hugo K.  
APPLICANT: Courtney-Guterson, Neal  
APPLICANT: Keller, Janis  
APPLICANT: Nijjar, Charanjit S.  
APPLICANT: Ralston, Edward J.  
TITLE OF INVENTION: PH Genes and Their Uses  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/537,715  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/049,282  
FILING DATE: 16-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US94/04173  
FILING DATE: 15-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 012176-003410US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 383 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-537-715-4

Query Match 33.7% Score 43.5; DB 2; Length 383;  
Best Local Similarity 52.4% Pred. No. 78;  
Matches 11; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

QY 4 LEALADQTDALQSEEAAYVK 23  
||| I:|:| I:|:| I:|:| I:|:|  
DB 244 LEARNQTEATLQTKDTGTVK 264

RESULT 13  
PCT-US94-04173-4  
Sequence 4, Application PC/TUS9404173  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: PH GENES AND THEIR USES  
NUMBER OF SEQUENCES: 4  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04173  
FILING DATE: 15-APR-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/049,282  
FILING DATE: 16-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 12176-34-1PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 383 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-04173-4

Query Match 33.7% Score 43.5; DB 5; Length 383;  
Best Local Similarity 52.4% Pred. No. 78;  
Matches 11; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

QY 4 LEALADQTDALQSEEAAYVK 23  
||| I:|:| I:|:| I:|:| I:|:|  
DB 244 LEARNQTEATLQTKDTGTVK 264

RESULT 14  
US-08-657-392-2  
Sequence 2, Application US/08657392  
Patent No. 5843634  
GENERAL INFORMATION:  
APPLICANT: Brate, E.M.  
APPLICANT: Brennan, C.A.  
APPLICANT: Bridon, D.P.  
APPLICANT: Jaffe, K.D.  
APPLICANT: Krafft, G.A.  
APPLICANT: Mandeckl, W.

7 AFFILIANT.  
7  
7 TITLE OF INVENTION: PH GENES AND THEIR USES  
7  
7 NUMBER OF SEQUENCES: 4  
7  
7 COMPUTER READABLE FORM:  
7  
7 MEDIUM TYPE: Floppy disk  
7  
7 COMPUTER: IBM PC compatible  
7

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/320,774  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA: US 08/762,106  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brotman, Harris F.  
REGISTRATION NUMBER: 35,461  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 654-2428  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 527 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: Internal  
US-09-320-774-9

Query Match 34.5%; Score 44.5; DB 4; Length 527;  
Best Local Similarity 51.7%; Pred. No. 82;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

QY 1 SPALEALADDTALQSEEAAYVK--ADNA 27  
:||| | | | :||| | | |  
Db 380 ADAVEAAAEELTA-QAEAAAKWEADKA 407

## RESULT 6

US-08-745-404-3  
Sequence 3, Application US/08745404B  
Patent No. 6096717

GENERAL INFORMATION:  
APPLICANT: Jarvik, Jonathan W.  
TITLE OF INVENTION: Method For Producing Tagged Genes,  
FILE OF INVENTION: Transcripts And Proteins  
FILE REFERENCE: 2087-961422  
CURRENT APPLICATION NUMBER: US/08/745,404B  
CURRENT FILING DATE: 1996-11-08  
EARLIER APPLICATION NUMBER: 08/000,619  
EARLIER FILING DATE: 1993-01-05  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3

LENGTH: 552  
TYPE: PRT  
ORGANISM: Chlamydomonas  
US-08-745-404-3

Query Match 34.5%; Score 44.5; DB 3; Length 552;  
Best Local Similarity 51.7%; Pred. No. 87;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

QY 1 SPALEALADDTALQSEEAAYVK--ADNA 27  
:||| | | | :||| | | |  
Db 405 ADAVEAAAEELTA-QAEAAAKWEADKA 432

## RESULT 7

US-08-961-083-192  
Sequence 192, Application US/08961083  
Patent No. 6159459

GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue

CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 192:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 344 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-083-192

Query Match 34.1%; Score 44; DB 4; Length 344;  
Best Local Similarity 55.0%; Pred. No. 58;  
Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 8 ADOTDALQSEEAAYVKADNA 27  
| | | | | | | | | | | | | | |  
Db 3 AGOTDASQIEKAAYVSGGKA 22

## RESULT 8

US-08-960-022-14  
Sequence 14, Application US/08960022  
Patent No. 5976837

GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallie, Edward R.  
APPLICANT: Racie, Lisa A.  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/960,022  
FILING DATE:  
CLASSIFICATION: 514

; Sequence 2, Application US/08745404B  
; Patent No. 6096717  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: Method For Producing Tagged Genes,  
; TITLE OF INVENTION: Transcripts And Proteins  
; FILE REFERENCE: 2087-961422  
; CURRENT APPLICATION NUMBER: US/08/745.404B  
; CURRENT FILING DATE: 1996-11-08  
; EARLIER APPLICATION NUMBER: 08/000.619  
; EARLIER FILING DATE: 1993-01-05  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 516  
; TYPE: PRT  
; ORGANISM: Chlamydomonas  
US-08-745-404-2

Query Match 34.5%; Score 44.5; DB 3; Length 516;  
Best Local Similarity 51.7%; Pred. No. 80;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Qy 1 SDALEALADQTDALQSEEAAYVK--ADNA 27  
Db 369 ADAVEAAAEELTA-QAEAAANAKWEADKA 396

RESULT 3  
US-09-320-774-8  
; Sequence 8, Application US/09320774  
; Patent No. 6265545  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
; TITLE OF INVENTION: TAGGING  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harris Brotman  
; STREET: 202 Coast Blvd., Suite 111  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/320.774  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/762.106  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brotman, Harris F.  
; REGISTRATION NUMBER: 35,461  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 654-2428  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-09-320-774-8

Query Match 34.5%; Score 44.5; DB 4; Length 516;  
Best Local Similarity 51.7%; Pred. No. 80;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Qy 1 SDALEALADQTDALQSEEAAYVK--ADNA 27  
Db 369 ADAVEAAAEELTA-QAEAAANAKWEADKA 396

RESULT 4  
US-08-762-106-9  
; Sequence 9, Application US/08762106  
; Patent No. 5948677  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
; TITLE OF INVENTION: TAGGING  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harris Brotman  
; STREET: 202 Coast Blvd., Suite 111  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/762.106  
; FILING DATE: 09-DEC-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brotman, Harris F.  
; REGISTRATION NUMBER: 35,461  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 654-2428  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 527 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-762-106-9

Query Match 34.5%; Score 44.5; DB 2; Length 527;  
Best Local Similarity 51.7%; Pred. No. 82;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Qy 1 SDALEALADQTDALQSEEAAYVK--ADNA 27  
Db 380 ADAVEAAAEELTA-QAEAAANAKWEADKA 407

RESULT 5  
US-09-320-774-9  
; Sequence 9, Application US/09320774  
; Patent No. 6265545  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
; TITLE OF INVENTION: TAGGING  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harris Brotman  
; STREET: 202 Coast Blvd., Suite 111  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 04:45:42 ; Search time 4.34225 Seconds  
(without alignments)  
157.503 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALEALADQTDALQSEAAVVKADNA 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44.5	34.5	516	2	US-08-762-106-8
2	44.5	34.5	516	3	US-08-745-404-2
3	44.5	34.5	516	4	US-09-320-774-8
4	44.5	34.5	527	2	US-08-762-106-9
5	44.5	34.5	527	4	US-09-320-774-9
6	44.5	34.5	552	3	US-08-745-404-3
7	44	34.1	344	4	US-08-961-083-192
8	44	34.1	514	2	US-08-960-022-14
9	43.5	33.7	382	1	US-08-049-2828-2
10	43.5	33.7	382	2	US-08-537-715-2
11	43.5	33.7	382	5	PCT-US94-04173-2
12	43.5	33.7	383	2	US-08-537-715-4
13	43.5	33.7	383	5	PCT-US94-04173-4
14	43.5	33.7	449	2	US-08-657-392-2
15	43.5	33.7	449	5	PCT-US94-02539-2
16	43.5	33.7	471	2	US-08-657-392-31
17	43.5	33.7	471	5	PCT-US94-02539-31
18	43.5	33.7	528	4	US-08-928-2138-8
19	43.5	33.7	711	1	US-08-235-838-7
20	43.5	33.7	711	2	US-08-465-4738-7
21	43	33.3	407	2	US-08-934-846-4
22	43	33.3	407	4	US-09-238-557-4
23	43	33.3	418	2	US-08-934-846-2
24	43	33.3	418	4	US-09-238-557-2
25	43	33.3	744	1	US-08-764-100-25
26	42.5	32.9	1611	2	US-08-804-227C-5
27	42	32.6	102	2	US-08-710-749-8

28 42 32.6 427 4 US-09-521-668B-18 Sequence 18, Appl  
29 42 32.6 427 4 US-09-521-668B-20 Sequence 20, Appl  
30 42 32.6 2890 4 US-09-413-814-67 Sequence 67, Appl  
31 42 32.6 3798 3 US-09-335-409-6 Sequence 6, Appl  
32 42 32.6 3798 4 US-09-568-102-6 Sequence 6, Appl  
33 42 32.6 3798 4 US-09-567-969-6 Sequence 6, Appl  
34 42 32.6 3798 4 US-09-568-480-6 Sequence 6, Appl  
35 42 32.6 3798 4 US-09-568-486-6 Sequence 6, Appl  
36 42 32.6 3798 4 US-09-568-472-6 Sequence 6, Appl  
37 41 31.8 173 3 US-08-669-408B-8 Sequence 8, Appl  
38 41 31.8 370 5 PCT-US93-01676A-4 Sequence 4, Appl  
39 41 31.8 395 1 US-07-876-280-9 Sequence 9, Appl  
40 41 31.8 395 1 US-08-049-783-6 Sequence 6, Appl  
41 41 31.8 395 1 US-08-316-301A-10 Sequence 10, Appl  
42 41 31.8 395 2 US-08-904-278-10 Sequence 10, Appl  
43 41 31.8 395 4 US-09-076-137-10 Sequence 10, Appl  
44 41 31.8 395 4 US-09-222-594-10 Sequence 10, Appl  
45 41 31.8 395 5 PCT-US92-03624-10 Sequence 10, Appl

## ALIGNMENTS

RESULT 1  
US-08-762-106-8  
; Sequence 8, Application US/08762106  
; Patent No. 5948677  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
; NUMBER OF INVENTIONS: TAGGING  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harris Brotman  
; STREET: 202 Coast Blvd., Suite 111  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/762,106  
; FILING DATE: 09-DEC-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brotman, Harris F.  
; REGISTRATION NUMBER: 35,461  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 654-2428  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-762-106-8

Query Match 34.5%; Score 44.5; DB 2: Length 516;  
Best Local Similarity 51.7%; Pred. No. 80;  
Matches 15: Conservative 4; Mismatches 7; Indels 3; Gaps 2;

QY 1 SDALEALADQTDALQSEAAVVK--ADNA 27

Db 369 ADAVEAAALTA-QAEAAARWEADKA 396

RESULT 2

US-08-745-404-2



```
/transl_table=11
/product="L-2,4-diaminobutyrate decarboxylase"
/protein_id="AG20873.1"
/db_xref="GI:10584266"
/translation="MEACADAVLDVIGGAHGYSGAYRDLRLDLADTTALPAEGHPL
EDVLAARDVLNANVHPDETCVAHLOCPPMVPALAAEALLTASNOSLDSFOAPAA
TVEERLIDLALVELGPAAGVVTGGGTESNHOALLARDDYVETVPGOSVREHGL
PAAQDLRLADAHFTAAQSAALLGIGEDAVYIPTDGRHRMDATALREVERLND
AKHPVAVGTAGTDFGSDLSAVDAVDAATHDVLVHVDAAFGMDALVSDORSLRS
GIDRASVADPHKFLFPQIACGALLVADGESFELMNAAYLNPAGDAPNLVAKST
RTTRRDALPVYAFRTLGRDIGALVDRTLALADVAGLLRADAFELACEPFLNAV
TFRYRERHAHICDQCPWADHVTEAREOLFQDAGTAVAVRTTVDRAHVKETLMNPRT
TVADIOSVLVAFKGAATIEAADRTAPITTAAGDQ"
4376..6199
/gene="lucA"
/notes="VNG6212G"
4376..6199
/gene="lucA"
/notes="lucA"
/codon_start=1
/transl_table=11
/product="iron transport protein A"
/protein_id="AAG20874.1"
/db_xref="GI:10584267"
/translation="MTGVCAVADTTPAAHAESATLHAFNLCLRETDAGEVVSRAVPD
GVPACQADRVRVPPFAQSTTVFAPLRYASATGRHLFPAVYAPVAPVAPAGATP
SFIDAGVASLARRELALDGDPTAGRDGLSERVLASQATERFVAARDTPROHERLYD
TVTSFQASQSLVYGHMHPTNSRTGIPPHQOSRYAPELRAGFQLRVPAADPALVET
MAAGDASATFELRSVVADATVPPESEATLVPVHPQADYLAHQPHVAAIAARDELUR
LGAFGPTFTPTSSVTLSPDPAPFMYKGSIAVEITNAERTNKRPELARGVAVSELIA
GFGDAPADFPFAIIEEDPAATLTLGDCAESGEFAVLRENPFAGDDTDTNVTVAALC
OQLDGPAAARIVRSIATREGRTESAVAREWFEQYLAHTVVRPVWLYIQOQVLEAH
QNTLVRLDDGRPTAGFYRDNQGYFFPASQAADVDNDLPGVCERADITVCPDAIDER
LRYIVFNAGVYVNALGAGGVVDERALLAVLRTELESIAEHEDPADSTLVSAALLTERR
VYKCANLITRFRERDELEADLETQSIVYTAIENPLVTRVDQ"
6196..6906
/gene="lucB"
/notes="VNG6213G"
6196..6906
/gene="lucB"
/notes="lucB"
/codon_start=1
/transl_table=11
/product="iron transport protein B"
/protein_id="AAG20875.1"
/db_xref="GI:10584268"
/translation="MTAPLPRITSDYAYRRPDQIDRTVAFRRVALDRDLRLHAWLT
SDHVTPTSLGAPLPAFHQAIGDHDHVTAYVGCIDGVPMSYVRYVWVAAADPLADH
YDADNDQGIHILIGPPEYLGCGTALLRAMTALAFRAHAATDRVVAEPDANDAAIA
AFERAGFEPRETFEPHEDKATRVCDRFRFVGAFSPCAIAANGPPADGPVANHDDA
THRCPSGEGHGNSTLEADR"
6903..8336
/gene="hxA"
/notes="VNG6214G"
6903..8336
/gene="hxA"
/notes="hxA"
/codon_start=1
/transl_table=11
/product="monooxygenase"
/protein_id="AAG20876.1"
/db_xref="GI:10584269"
/translation="MTESVHDLVGVGLGFPNFIGLAALIEGSDVDVPLDAALFLOKPEFV
WHDGMLEICTTLEVPFLADLVLTADNPYSYLNLYREDHRLYEFYFTFTGIPREY
NDYCWVWASLAAPFRRVVDITAIIDDATPANFRVTAVEPADARSATIDDDAQHQ
YKARLDVGVGRPVPPAFRDLPAQDVHFTAAVREHRAAVCDADSVAVIGSQSAAE
VYHDLITRSTSDRSPRVDMVTSAGFPMWYKLGLOHFTPEYTDYVDLQPVKDE
LIPNQDITLYKGDIPETSTRIYDTLYEQSIGRDPWGMVTAIEVGAFTPGGRYVD
CHRRDITGFATETDAVILGTGYHRPTGFLSSLEPLRLARDDDRLAVDREYGVGDL
PGCLFQVNAELHTGVGAPDLGAYRNAVILDRDLADAPYVDRDRTVPQDFAPEAYA
SHAPGGRVDDDPPTPPAAGQPNDRHQ"
8362..10188
/gene="lucC"
```

## CDS

```
/note="VNG6216G"
8362..10188
/gene="lucC"
/notes="lucC"
/codon_start=1
/transl_table=11
/product="iron transport protein C"
/protein_id="AAG20877.1"
/db_xref="GI:10584270"
/translation="MNTYEDTLDQDAVTDATWATVRSLLAKLLREFTYBELLAPERTT
DIDRGRYLVLPAAADATVRYRFAEERLMDTLRVFPOTIERADADSWATATDP
IRFLDRLDTELTLTGHLIREYRNTLLADAHIEARSAPESFADCSVAEYESMTG
HNKIFNKGRLGMDVEDYDAPEREQTTLIGNLAVRTDRVTEETVSGLTPTGLLGE
LGQDDDFMATLADRGLDPEQETLLPVHNYQWETTITISYASELADALVPLNGADS
YLPQGSVRTLFNVDAPEKHVKLPKIRNTLVWRGLPGDRTAAPVTEYITDIRDGD
PFLRDDCELVLFEVAGVNDHPAFSELEDNAYQYDELGAVMRESITDAIPDSQAV
TLAAILHEADGTAVISFERSPLSELAWLTELFLPLLLHLYLVRYGTVFSPHGEN
TIVLDEEVPTRLGVKDFVDVNIQDPLPELADLDADLRRLRSEPPNGLQCFYFC
GLFVGVRVYLSDLVTAHHGYDETEFTQVRDAITAVQAEFPELSDREFTDLLQPTFT
KICLNNRLTEEGYSDTASRPAAEHGTVTNALHEVAGDAE"
complement(10189..10842)
/gene="trkA3"
/notes="VNG6218G"
complement(10189..10842)
/gene="trkA3"
/notes="trkA3"
/codon_start=1
/transl_table=11
/product="Trk potassium uptake system protein"
/protein_id="AAG20878.1"
/db_xref="GI:10584271"
```

## Alignment Scores:

Prod. No.:	674	Length:	12423
Score:	51.00	Matches:	12
Percent Similarity:	65.38%	Conservative:	5
Best Local Similarity:	46.15%	Mismatches:	9
Query Match:	39.53%	Indels:	0
DB:	1	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AE005155 (1-12423)

QY 2 AspaLaleuGlualaleuAlaAspGlnThrAspaLaleuGlnSerGluGluAlaVal 21

Db 3981 GACGGACTCGGGCGGCTGCGTGCATCGAACGCTCGCGCTCGCGGCGGCGCTG 4040

QY 22 VallysAlaAspAsnAla 27

Db 4041 CTGCGTGCTGACCCGGCG 4058

## RESULT 40

AF072878

LOCUS

DEFINITION

Tetrahymena thermophila ciliary outer arm dynein beta heavy chain

ACCESSION

AF072878

VERSION

AF072878.2

KEYWORDS

SOURCE

ORGANISM

Tetrahymena thermophila

Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;

Hymenostomatida; Tetrahymenina; Tetrahymena.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

AF072878 14941 bp DNA linear INV 21-DEC-1999

Tetrahymena thermophila ciliary outer arm dynein beta heavy chain

(DYH4) gene, complete cds.

AF072878 1 bases 1 to 14941

AF072878.2 GI:5209335

Tetrahymena thermophila.

Tetrahymena thermophila

Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;

Hymenostomatida; Tetrahymenina; Tetrahymena.

Lee, S., Wisniewski, J. C., Dentler, W. L. and Asai, D. J.

Gene knockouts reveal separate functions for two cytoplasmic

dyneins in Tetrahymena thermophila

Mol. Biol. Cell 10 (3), 771-784 (1999)

99169020

10069817

2 (bases 1 to 14941)

Lincoln, L. M., Gibson, T. M., Asai, D. J. and Forney, J. D.

A gene knockout reveals that dynein beta heavy chain is required in

Tetrahymena thermophila







```

/chromosome="3R"
/map="84A-84B"
/clone="BACR32J03 (D1350)"
/clone_lib="RPC1-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACe3.6)"
BASE COUNT 55773 a 39290 c 38978 g 56601 t
ORIGIN

Alignment Scores:
Pred. No.: 1e+04 Length: 190642
Score: 51.50 Matches: 13
Percent Similarity: 65.38% Conservative: 4
Best local Similarity: 50.00% Mismatches: 8
Query Match: 39.92% Indels: 1
DB: 3 Gaps: 1

US-09-847-539A-6_COPY_59_86 (1-28) x AC095015 (1-190642)

Qy 2 AsphalateuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
|||||:::|||||::: ||||||| ||| ::: |||
Db 3240 GACGCTTTAAAGCGTTTCTTCACAGACTGAT---TTGAGCCATCCGATGAGTGATATT 3296

Qy 22 vallyalaaspanAla 27
|||||:::|||||
Db 3297 TTTAAGCGACACAATGCC 3314

RESULT 37
AE003673
LOCUS
DEFINITION
Drosophila melanogaster genomic scaffold 142000013386040 section 3
of 5, complete sequence.
ACCESSION
AE003673 AE002699
VERSION
AE003673.2 GI:10727114
KEYWORDS
HTG.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 309357)
Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Ananides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Rogers,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Watters,J.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews,Pfannkuch,C., Baldwin,D., Ballwe,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Bertram,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.A., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Fabios,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieta,S.,
Flusschmann,W., Fosler,C., Gabriellian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Idegvam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattai,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Klanos,I.,
Stimpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Swirskas,R., Tector,C., Turner,R.,

```

```

ACCESSION AC095014
VERSION AC095014.1 GI:15624856
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
DEFINITION Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 170801)
REFERENCE 1
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Unpublished
2 (bases 1 to 170801)
REFERENCE 2
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Direct Submission
Submitted (15-SEP-2001) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu.
Location/Qualifiers
1. 170801
/organism="Drosophila melanogaster"
/strain="v: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="83F-84A"
/clone.lib="BAC19D23 (D1349)"
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.6"
BASE COUNT 49244 a 35184 c 35797 g 50576 t
ORIGIN
Alignment Scores:
Pred. No.: 8.91e+03 Length: 170801
Score: 51.50 Matches: 13
Percent Similarity: 65.38% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 39.92% Indels: 1
DB: 3 Caps: 1

```

---

```

US-09-847-539A-6_COPY_59_86 (1-28) x AC095014 (1-170801)
Qy 2 AspaLaLeuGluaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGlUGluAlaLaVal 11
|||||:|||||::: |||||:|||||:::
Db 170223 GACGCTTTAAAGCGCTTTCTTCACAGACTGAT---TTGAGCCATCCGATGAGTGATATT 170279
Qy 22 VallysAlaAspAspAla 27
|||||:|||||:|||||:|||||:|||||:
Db 170280 TTTAAGGCAGACAAATGCC 170297
RESULT 36
AC095015
LOCUS Drosophila melanogaster, chromosome 3R, region 84A-84B, BAC clone
DEFINITION BACR3J03, complete sequence.
AC095015
VERSION AC095015.1 GI:15624857
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
DEFINITION Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 190642)
REFERENCE 1
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Unpublished
2 (bases 1 to 190642)
REFERENCE 2
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Direct Submission
Submitted (15-SEP-2001) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu.
Location/Qualifiers
1. 190642
/organism="Drosophila melanogaster"
/strain="v: cn bw sp"
/db_xref="taxon:7227"

```

/db\_xref="GI:1457550"  
/translation="MSDVTDAVALVDHGTPVPALVLASLAALATATAGFHTWRSHHG  
CAPPADTGRAPPTAGTGPAGSGGSGGAGFSAALWRMTTVDAPGSL  
AVLCAALADRLDILSLQTHPLTCTVDFLLRAPGELSAELAENVYTAGGARTWTE  
RADHLDVADTRVLGLAARVALDAELPLALRQLLGRCTIRSLPAPADDDGAGARAA  
ADPEGALEDVTLRLRDEGVIISVERPPLPFTFAEFARARALVELDARLPRVPRGR  
DVLTLPEGSDDTVRRADTRDYPAAKAMHERGSARTLGMRYHGPVGDADRYLHLLSPR  
FGTLAVTASGRVLGLHLLMDGDETEVALLVEDGMQRGIGGELLARLVAMAADAG  
CESVVAITQASNTGMVAAMRGLSLPLDYIEEGLVLTARLDPAQAARALPQAEIRG  
RE"

## misc\_feature

5557..5766  
/gene="SCBAC36F5.04"  
/note="Pfam match to entry PF01842 ACT, ACT domain, score  
17.60, E-value 0.29"  
6310..6516  
/gene="SCBAC36F5.04"  
/note="Pfam match to entry PF00583 Acetyltransf,  
Acetyltransferase (GNAT) family, score 36.20, E-value  
7.7e-07"  
complement(6676..7899) .  
/gene="SCBAC36F5.05c"  
complement(6676..7899)  
/gene="SCBAC36F5.05c"  
/note="SCBAC36F5.05c, possible lyase, len: 407 aa; similar  
to SM:MTB\_HAEIN (EMBL:U32694) Haemophilus influenzae  
cystathionine gamma-synthase (EC 4.2.99.9) MetB or HI0086,  
369 aa; fasta scores: opt: 625 Z-score: 685.3 bits: 135.6  
E(): 1.5e-30; 34.127% identity in 378 aa overlap. Contains  
Pfam match to entry PF01053 Cys\_Met\_PP, Cys/Met  
metabolism PLP-dependent enzyme and match to Prosite entry  
PS01050 Uncharacterized protein family UPF0031 signature  
2".

## gene

complement(6676..7899) .

## CDS

/codon\_start=1  
/transl\_table=11  
/product="putative lyase"  
/protein\_id="CAC42841.1"  
/db\_xref="GI:1457551"  
/translation="MDSARTHTSPDDVRFETAPAGRALATRAVHAGRDDLLAAQGLHAP  
PIDLSTYPSYDSRGEAARIDAFATGAEPDGPVYGRGNPTVARFETALARLEGTE  
ADAFASGMAALSALLVLRGSMGLRHVAVRPLVYCGSDHLLTAGLLGSEVTTWTPAGV  
AADLRPTDGLVLVESPNPTLAEDVLRAVHACGVPVLLVDNTATPVLRPAEHRAR  
LVLSATYLVGGHGDVLAVGVACDEEFAGRLRQVRFPATGGVHLPLAGVLLRLGLSTLP

## Alignment Scores:

Pred. No.: 1  
Score: 1.55e+03 Length: 32608  
Percent Similarity: 51.50 Matches: 17  
Best Local Similarity: 50.00% Conservative: 3  
Query Match: 42.50% Mismatches: 7  
DB: 39.92% Indels: 13  
Gaps: 2

US-09-847-539A-6\_COPY\_59\_86 (1-28) x SCBAC36F5 (1-32608)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThr----- 11  
Db 13175 GCGGAGCGCGCGCGCGCGCGTCTCGTGCAGACACCGCCGAGTGTCAACGACGGCTC 13116  
QY. 12 AspAlaLeuGlnSerGluGluAlaAlaVal-----ValLysAlaAspAsnAla 27  
Db 13115 GACGCTTGAAGGCGGAGCGCGCGTCTCGTGCAGACACCGCCGAGTGTCAACGACGGCTC 13056

## RESULT 33

AC012724  
LOCUS AC012724.1 35031 bp DNA linear HTG 03-NOV-1999  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
pieces  
ACCESSION AC012724  
VERSION AC012724.1 GI:6223583  
KEYWORDS HTG; HTGS-PHASE2.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 35031)  
AUTHORS Adams M. and Venter J.C.  
TITLE Direct Submission  
JOURNAL Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
COMMENT This sequence was identified as CDM:10209782 by the submitter.  
For further information on this sequence you may e-mail to  
fly@celera.com.

\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

## FEATURES

1..35031  
Location/Qualifiers  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"

BASE COUNT 9943 a 7273 c 7242 g 10573 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1  
Score: 1.67e+03 Length: 35031  
Percent Similarity: 51.50 Matches: 13  
Best Local Similarity: 65.38% Conservative: 4  
Query Match: 50.00% Mismatches: 8  
DB: 39.92% Indels: 1  
Gaps: 2

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC012724 (1-35031)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21  
Db 13803 GACGCTTTAAAGCGTTTCTTCACAGACTGAT---TTGAGCCATCCGATGATGATAT 13859

QY 22 ValLysAlaAspAsnAla 27

Db 13860 TTTAAGGCGACAAATGCC 13877

## RESULT 34

AE001572.2/c

## WPCOMMENT

Sequence split into 5 fragments LOCUS AE001572 Accession AE001572

Fragment Name Begin End  
AE001572\_0 1 110000  
AE001572\_1 100001 210000  
AE001572\_2 200001 310000  
AE001572\_3 300001 410000  
AE001572\_4 400001 429825

Continuation (3 of 5) of AE001572 from base 200001 (AE001572 Drosophila melanogaster

## Alignment Scores:

Pred. No.: 5  
Score: 5.6e+03 Length: 110000  
Percent Similarity: 51.50 Matches: 13  
Best Local Similarity: 65.38% Conservative: 4  
Query Match: 50.00% Mismatches: 8  
DB: 39.92% Indels: 1  
Gaps: 3

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AE001572.2 (1-110000)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21

Db 20551 GACGCTTTAAAGCGTTTCTTCACAGACTGAT---TTGAGCCATCCGATGATGATAT 20495

QY 22 ValLysAlaAspAsnAla 27

Db 20494 TTTAAGGCGACAAATGCC 20477

## RESULT 35

AC095014

## LOCUS

DEFINITION AC095014  
Drosophila melanogaster, chromosome 3R, region 83F-84A, BAC clone  
BACH19D23, complete sequence.  
170801 bp DNA linear INV 15-SEP-2001

Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

1 (bases 1 to 32608)  
 Redenbach, M., Kieser, H.M., Denapalite, D., Eichner, A., Cullum, J., Kinashi, H. and Hopwood, D.A.  
 A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
 Mol. Microbiol. 21 (1), 77-96 (1996)

2 (bases 1 to 32608)  
 Saunders, D. and Harris, D.  
 Unpublished  
 Direct Submission  
 Cerdano, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.  
 Submitted (27-JUN-2001) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:  
 Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
 {URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>} CDS are numbered using the following system eg SC7B7.0lc. SC (S. coelicolor), 7b7 (cosmid name), .0l (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid BAC36F5 overlaps cosmid 2StG18 and cosmid 2StG61.  
 Location/Qualifiers  
 1. .32608  
 /organism="Streptomyces coelicolor"  
 /db\_xref="taxon:1902"  
 1. .99  
 /gene="SCBAC36F5.01"  
 /note="nominal overlap with Streptomyces coelicolor cosmid 2StG18"

misc\_feature  
 1. .32608  
 /organism="Streptomyces coelicolor A3(2)"  
 /strain="A3(2)"  
 /db\_xref="taxon:100226"  
 /clone="cosmid BAC36F5"  
 1. .891  
 /gene="SCBAC36F5.01"  
 <1. .891  
 /gene="SCBAC36F5.01"  
 /note="SCBAC36F5.01, possible secreted phosphatase (fragment), len: >296 aa; similar to C-terminal region of SW:PPBD\_BACSU (EMBL:U49060) Bacillus subtilis alkaline phosphatase D precursor (EC 3.1.3.1) PhOD, 556 aa; fasta scores: opt: 277 Z-score: 319.4 bits: 68.0 E(): 3.5e-10;

30.258% identity in 271 aa overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative secreted phosphatase (fragment)"  
 /protein\_id="CAC42837.1"  
 /db\_xref="GI:1457547"  
 /translation="LDENLRFAQVPSIVQDDHEVRNNYPGQVLDYDDRYTEKRVDVLAARARAFAPYPPVSLRRCAREVRHVLHOGPLLDVFLDMRTYRNANSPGEQAVDPQGLRGQLEWLKRELASRAVWKVIADMPGLGLVYPTTTEGAPHEIAAQSDPGAPLGRLEQIAELLRFVKHRRITGTWLTADVHHTSAQHYQPSRAAFGDFPEPWFVSGPLSAGAFPASBLDGTGTFGPFRVFKAPTASVSPAGGYQFFGEVDIDGSGTAENTVRLREQDGTVLFTKVLRPQVQGO"  
 1042. .2445  
 /gene="SCBAC36F5.02"  
 1042. .2445  
 /gene="SCBAC36F5.02"  
 /note="SCBAC36F5.02, unknown, len: 467 aa"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein SCBAC36F5.02"  
 /protein\_id="CAC42838.1"  
 /db\_xref="GI:1457548"  
 /translation="MTRSLQAVAYRRPSVLESAGGGHGLGETSRGATPAGAVDHPRFAGLDTAPQKASTALLAVADVAARYQQQLRASLDPPVTGSDRLRFESFGCGGYVARDLVEAGLDGGEVGHGTTNDVNNPLREALSRIGADDPLHLRVGPELAVTTLDGPVVKKVPPLDMRLRGPFAFOVIAAGFDLRAFLDAAEAVRFLRALPGRGRSAATLRWVVPAGRSRTPTRAVPGAVCLPGPERLVALQRLRHATALRVYGPVAVASGAATASAAVVLTPGMRLTLLSPDAARGEGGGLDALADEAADAELISVLLAWEPIDRTADLA AASGLTAERVRAALVRIGTSGRVGYDTAEAYFHRELPHYDAERYERINRLRSARLV AAGGVLDGPIGTVTATDGHVHRVREAGVLSCLMWAKYRGGRGPCKHALLAVRMVR RGSQSGAVRIDGGAR"  
 2454. .5177  
 /gene="SCBAC36F5.03"  
 2454. .5177  
 /gene="SCBAC36F5.03"  
 /note="SCBAC36F5.03, unknown, len: 907 aa"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein SCBAC36F5.03"  
 /protein\_id="CAC42839.1"  
 /db\_xref="GI:1457549"  
 /translation="MASAVATPAATATATATDTTGLVAARAGTARVALLDGMTOA ORRAAPELKAVRQELRTARNSDSRRAHPALHAAGACOTGAANANLAAVDMRWNS QASPAVLLHVLGDRCWLDVAHRAQPAASNVSELMAGLVRLSGCPVPPTTAVV QGWLTVGGQWQGGCTVCDRLRRDPLHAEVLVALFRTEGVTQLGWSGDPSPSWTSA LARLTGEGVLDRRSVVDACVARLLRGAPAEIRVFLRLKDLALTREERERRAADNTA LKRSAAAGELLPAARAFGHEDSDAQERLAKLVVERHVKTAAEVREELAGAAEQLLP ALRAAVRTLCATPPDSVPVAHEEVLPPQPPVRLAPAPAAVVAEEISALLASGSD VAVFERALDGLVRHAILDRDALLAEPLPVVARKWMAEGEPHGRSDDFPARRHELFDDARAFDILLATLYGKVRTDTLIRVYHDLGNSGCAINSLARPEARVWEVNAIRLTPD QPFLLATPTWTGLLEPGELVDLRADYRRAGARVSAVDFGQALLRVDNRNPAALAA ERAAALTEGERLARLLTERAAPSVTRTADTRILVELGEVPLDRAEFPFPAALRLL GRPVNPDSDHWYCSHNSGAMRPHWIFALLPERPELVAARLLRDVSEAAVHDDQGTAAAV LPHLADADGPAEAGAVHL5VAYGLGARHAEADRLAVALDALLVLAARGRLAEGRGLADICQ LVRRGAVKPARLADAVRTAAATCANATVWAVLROVLPVLLADLSTGCATTTSSARGLE LLVAAAEACBRTADEGQSLPHLSGVADRRGTSLRVLTQARRLEALAAAPAAA"  
 5293. .6651  
 /gene="SCBAC36F5.04"  
 5293. .6651  
 /gene="SCBAC36F5.04"  
 /note="SCBAC36F5.04, hypothetical protein, len: 452 aa; C-terminal part similar to TR:09FC64 (EMBL:AL391515) Streptomyces coelicolor putative acetyltransferase SC4B10.23, 163 aa; fasta scores: opt: 287 Z-score: 300.0 bits: 63.3 E(): 4.2e-09; 42.857% identity in 133 aa overlap. Contains Pfam match to entry PF01842 ACT, ACT domain and PF00593 Acetyltransf, Acetyltransferase (GNAT) family"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAC42840.1"

```

DB:          33          Gaps:          2

US-09-847-539A-6_COPY_59_86 (1-28) x AC056141 (1-850)

Qy  1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAla-----LeuGlnSerGlu 17
   ::::|||||:::  ::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  848 GCAGATGCGATCGAGGAATTCGCCACCCACCGCAGCGCGCCGACGAGTTCAGCAGGAG 789

Qy  18 GluAlaAlaValVallys-----AlaAspAsnAlaAla 28
   |||||:::  :::  |||||:::|||||
Db  788 CTTGACGCGCTCGCGCGCGACGACGCGGAGACGCGAGCG 750

RESULT 30
LOCUS      ATUIS22A                2888 bp    DNA        linear    BCT 26-APR-1993
DEFINITION Agrobacterium tumefaciens insertion element IS1131.
ACCESSION M82888
VERSION    M82888.1 GI:142231
KEYWORDS   Insertion element IS1131; transposase.
SOURCE     Agrobacterium tumefaciens
ORGANISM   Agrobacterium tumefaciens
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE  1 (bases 1 to 2888)
AUTHORS    Wabiko,H.
TITLE      Sequence analysis of an insertion element, IS1131, isolated from the
            nopaline-type Ti plasmid of Agrobacterium tumefaciens
JOURNAL    Gene 114, 229-233 (1992)
MEDLINE    92290280
FEATURES   Location/Qualifiers
            source          1..2888
                        /organism="Agrobacterium tumefaciens"
                        /db_xref="taxon:358"
                        156..164
            misc_feature    632 a 794 c 882 g 580 t

BASE COUNT  632 a 794 c 882 g 580 t
ORIGIN
Alignment Scores:
Pred. No.:      120          Length:      2888
Score:          51.50       Matches:      13
Percent Similarity: 72.00%   Conservative: 5
Best Local Similarity: 52.00% Mismatches:   6
Query Match:    39.92%      Indels:     1
DB:             1          Gaps:         1

US-09-847-539A-6_COPY_59_86 (1-28) x ATUIS22A (1-2888)

Qy  3 AlaLeuGluAlaLeu---AlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  975 GCTTTGCGCGACTGTCGCCGACAGCGCGGAGCTTGAGAGCGGAGGAGCGGAGTC 1034

Qy  22 ValLysAlaAspAsn 26
   ::::|||||  |||||
Db  1035 ATCAGCGGAGACTCC 1049

RESULT 31
AC001654
LOCUS      Drosophila melanogaster 12976 bp    DNA        linear    INV 21-JUL-1998
DEFINITION Drosophila melanogaster DNA sequence (P1 DS00189 (D16)), complete
ACCESSION AC001654 L48712 L39780 L39782 L39781 L39783
VERSION    AC001654.1 GI:2342707
KEYWORDS   HTG.
SOURCE     Drosophila melanogaster (Subclones in sac from P1 clone DS00189
            (D16)) DNA.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 12976)
AUTHORS    Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.I., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., PUNCH,E., Snir,E., Twomey,B., Wan,X.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
            Sequencing of antennapedia complex, homeotic genes
            unpublished (1997)
            2 (bases 1 to 12976)
            Martin,C.H., Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A.,
            Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
            Houston,K.A., Jaklevic,M.A., Kadner,K.E., Kim,K., Kim,S.F.,
            Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomotan,M.A.,
            Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
            Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C. and Palazzolo,M.J.
            Direct Submission
            Submitted (22-APR-1997)
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@hgc.lbl.gov.
            Library location: 136-2.
FEATURES   Location/Qualifiers
            source          1..12976
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /chromosome="3R"
                        /map="84A6-84B2"
                        /clone="P1 DS00189 (D16)"
                        /note="This sequence has not changed since its original
                        submission on 04/22/1997. It was resubmitted in order to
                        include all secondary accession numbers for the subclones
                        belonging to this clone."
BASE COUNT  3670 a 2713 c 2731 g 3862 t
ORIGIN
Alignment Scores:
Pred. No.:      587          Length:      12976
Score:          51.50       Matches:      13
Percent Similarity: 65.38%   Conservative: 4
Best Local Similarity: 50.00% Mismatches:   8
Query Match:    39.92%      Indels:     1
DB:             3          Gaps:         1

US-09-847-539A-6_COPY_59_86 (1-28) x AC001654 (1-12976)

Qy  2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21
   |||||  |||||  |||||  :::  |||||  |||||  |||||  |||||  |||||
Db  8074 GACGCTTTAAAAGCGTTTCTTCACAGACTGAT---TTGAGCCATCCGATGAGTGATATT 8130

Qy  22 ValLysAlaAspAsnAla 27
   |||||  |||||  |||||
Db  8131 TTTAAGGCAGACAATGCC 8148

RESULT 32
SCBAC36F5/c
LOCUS      SCBAC36F5                32608 bp    DNA        linear    BCT 28-JUN-2001
DEFINITION Streptomyces coelicolor cosmid BAC36F5.
ACCESSION AL592292
VERSION    AL592292.1 GI:14575546
KEYWORDS   AsnC-family transcriptional regulatory protein; cation efflux
            system protein; exonuclease; hydrolase; integral membrane protein;
            lyase; regulatory protein; secreted phosphatase; siderophore
            binding protein; sugar acetyltransferase; transcriptional
            regulatory protein.
SOURCE     Streptomyces coelicolor.
ORGANISM   Streptomyces coelicolor
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```





```

Oy      4  LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
|||||TTTT|  |||  :::::|||||
Db 246775  CTGGAAGCTCTGTTCAAGCAGGTGAGTCTCTCCAAAGCCGCGAGTTGCTCTCTTGCGCA 246834

Oy      24  AlaAspAsn 26
|||||TTTT|
Db 246835  GCCGACAAT 246843

RESULT 28
RME603646/c 303100 bp  DNA  linear  BCT 16-AUG-2001
LOCUS Rhizobium meliloti (Sinorhizobium meliloti) 1021 complete pSymB;
DEFINITION segment 5/6.
ACCESSION AL603646 AL591985
VERSION AL603646.1 GI:15140973
KEYWORDS
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
Finan,T.M., Weidner,S., Wong,K., Buhrmester,J., Chain,P., Gouzy,J.,
Vorholter,F.J., Hernandez-Lucas,I., Becker,A., Cowie,A., Gouzy,J.,
Golding,B. and Fuhler,A.
From the Cover: The complete sequence of the 1,683-kb pSymB
megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti
Proceedings of the National Academy of Sciences of the United
States of America. 98 (17), 9889-9894 (2001)
11481431
pubmed ahead of print
REMARK 2 (bases 1 to 303100)
REFERENCE Weidner,S.
AUTHORS Direct Submission
TITLE Submitted (07-JUN-2001) Weidner S., Universitaet Bielefeld,
JOURNAL Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld,
Germany
COMMENT Submitted on behalf of Universitaet Bielefeld, Biologie IV
(Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany and
Department of Biology, McMaster University, 1280 Main Street West,
Hamilton, Ontario, L8S 4K1 Canada
mailto:Stefan.Weidner@genetik.uni-bielefeld.DE
PEXO, pSymB.
FEATURES
source
join(1..134869,join(136189..275356,276016..303100))
/organism="Sinorhizobium meliloti"
/plasmid="pSymB"
/strain="1021"
/db_xref="taxon:382"
/focus
complement(167..967)
/gene="thid OR Smb20962"
complement(167..967)
/gene="thid OR Smb20962"
/EC-number="2.7.4.7"
/function="Small Molecule Metabolism; Biosynthesis of
cofactors, carriers; thiamin"
/note="Product confidence : probable
Gene name confidence : putative"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="probable phosphomethylpyrimidine kinase protein"
/protein_id="CAC49487.1"
/db_xref="GI:15140974"
/translation="MTAIALSIAGSDSGGAGIADLKTFPSALGVYASVITATAQN
TRGVATVEDVSAEIVSQMDAVFSLDVKVYKIGMVSRRETIAAIADGLRRFGKRVV
DPNVATSGDALRPDAVALIELLPALVYTPNLAEALMTGRAIAGDAENARQA
EAMRTGAHVLYVRGHLKGQEAIDLFDDGTLVRLPAGRIETRNDRHGTCTLSAALA
AGLAKGVPLIEAASAAKAYLHAASIAADRLEIGQGRGPVHHFHWKWD"
1002..1307
/gene="Smb20963"
1002..1307
/gene="Smb20963"

```

---

```

/function="MISCELLANEOUS; Not classified regulator"
/note="Product confidence : putative"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcriptional regulator protein"
/protein_id="CAC49488.1"
/db_xref="GI:15140975"
/translation="MRVROWTSSIDLTQTKHHGDPKGDEKANILPLAAIFPFLRRHDP
DOVGSGSTPSOHLVLPSECAIYAAVRCORAKACKTYLLGQPAKIYPOIGKL"
1412..2071
/gene="Smb20964"
1412..2071
/gene="Smb20964"
/function="CELL PROCESSES; Protection responses"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative anti-oxidant protein, AhpCTSA family"
/protein_id="CAC49489.1"
/db_xref="GI:15140976"
/translation="MSLRINDTAPDFTAETQTQNTINHENIGDGNVLFSPKNTPTV
CTTELGAMAGIEPEFRKRGYKIIIGISYDVPVESHKWNKIDKVAATGFEVDYPLIGDRDL
KVAKLYDMLPAGAGETSEGRTPADNATVRSYVIGPKKIKLILTYPMTTGRNFNEIL
RAIDSQILTAKHOVATPANQOGEDVLIITAAVSNEDAVQRFSGSDTVLPVLRKTKQPT
A"
complement(2189..3187)
/gene="Smb20965"
complement(2189..3187)
/gene="Smb20965"
/function="MISCELLANEOUS"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative exported enzyme, similar to
5-aminohexanoate-dimer hydrolase (EC 3.5.1.46) protein"
/protein_id="CAC49490.1"
/db_xref="GI:15140977"
/translation="MLVITLLALGFHSAADQSVTTLTLEELREELKPLKTVVY
ARDGETAEKGYRGHTPESTNISKASKSIVSALVGIAIDKGLDGEQKTAPLLKED
LPSQADPRINDITIGNLLSMQAGLRMSGPNYGRWVRNVRALAQPFDEPGGRM
LYSTASTHLLSAILTKVGGKTLALAREMLDPVEGFRIGAWERDQCILYLCGNQMAS
ARSLATGELYRNRCRTADGQVVPADVVELSWRPRTASRTGCGYGYGFWPTRIGDE
DVYAWYGGOMLYIVPSLNLTVVMTSDSGPSARNGYRDALHGVLAELIISTAKAD"
complement(3296..5236)
/gene="lac22 OR Smb20966"
complement(3296..5236)
/gene="lac22 OR Smb20966"
/EC-number="3.2.1.23"
/function="Small Molecule Metabolism; Degradation; carbon
compounds"
/note="Product confidence : putative
Gene name confidence : putative"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative beta-galactosidase protein"
/protein_id="CAC49491.1"
/db_xref="GI:15140978"
/translation="MCYYPEQNPGRKWAEDARRMVELGLSHVRIGEEFAWAKIEPRSGE
FHWLDEADIVIGKAGIKVITLPTTAPPAWLVNRYPILPLVDATGAKRFGARRHY
FCSRRYRSEAAIRTEAMARYGEHYIYHAWQTDNEYGDHDTIYSYSAEYAGAPRLML
AERYGSIDELNRAMGTSFWSMRYDSFEIDLPNNLVEEPSPTHGVDIFRFSDDQVKSF
NKAQVDIIRASHPCRPVTHNFMSTQDTHYRVGDEDIASWDVYPMGGLNGLAAR
DKRYLRVGDDQPAFNHDLRYAVGRGVVMEQPGPVNNAAHNAPDAGMVRWLTW
LAYAHGDMVSYFEWRQAPAEQOFHAGLLLPNSEADQGYLEVAEVAEMKRLPEGEV
RGAQVAIVLDYESTRWATRVLPQGRSYSASAVADWTYVARLVGDVDVDFIQHSDIAG
YKLILAPDLVIAEEAFVETRADAKVYFGARSGSKTRDMHIPGLPPGPAKLIDIS
VSRVESLPFHSFETVLYGNEAYEAGGWRETVRTSETVIASFDEYRNGALVGSDKA

```





\* the accession number will be preserved.

```

FEATURES
  source
    Location/Qualifiers
      1..36993
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
BASE COUNT  9781 a 8675 c 8725 g 9812 t
ORIGIN
Alignment Scores:
  Pred. No.:      1.48e+03      Length:      36993
  Score:          52.00         Matches:    11
  Percent Similarity: 65.22%    Conservative: 4
  Best Local Similarity: 47.83%  Mismatches: 8
  Query Match:     40.31%      Indels:    0
  DB:              2          Gaps:      0
US-09-847-539A-6_COPY_59_86 (1-28) x AC018007 (1-36993)
QY  4  LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db  30206 CTGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAGCGCCGAGTCTCTCTGTGTGGCA 30265
QY  24  AlaAspAsn 26
Db  30266 GCCGACAAT 30274
RESULT 25
AC011910
LOCUS
  Drosophila melanogaster clone RPC198-2017, *** SEQUENCING IN
  PROGRESS ***, 39 unordered pieces.
AC011910
VERSION 1
KEYWORDS AC011910.1 GI:6056126
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 133155)
  Muzny,D.M., Adams,C., Bailey,M., Barberis,J., Blankenburg,K.,
  Bodota,B., Bouck,J., Bowler,S., Brooks,A., Buhay,C., Bunak,C.,
  Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
  David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
  Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
  Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
  Guevara-Wesley,J., Harris,K., Hernandez,J., Hodgson,A., Hogue,M.,
  Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
  Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
  Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
  Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
  Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,
  Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
  Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,
  Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R., Tabor,P.,
  Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wabwah,M., Watlington,S.,
  Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
  Wrenford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
  Direct Submission
  2 (bases 1 to 133155)
  Unpublished
  Worley,K.C.
  Direct Submission
  Submitted (16-OCT-1999) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  TITLE
  JOURNAL
  AUTHORS
  COMMENT
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 39 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence.
    * As soon as it is available and the accession number will
  
```

```

* be preserved.
  1
  854: contig of 854 bp in length
  1695: contig of 841 bp in length
  2535: contig of 840 bp in length
  3358: contig of 823 bp in length
  4171: contig of 813 bp in length
  5125: contig of 954 bp in length
  6105: contig of 980 bp in length
  6952: contig of 847 bp in length
  7603: contig of 651 bp in length
  8573: contig of 970 bp in length
  9724: contig of 1151 bp in length
  10971: contig of 1247 bp in length
  12519: contig of 1548 bp in length
  14371: contig of 1852 bp in length
  15284: contig of 893 bp in length
  16905: contig of 1641 bp in length
  19536: contig of 2631 bp in length
  21604: contig of 2068 bp in length
  22985: contig of 1381 bp in length
  24383: contig of 1378 bp in length
  26424: contig of 2061 bp in length
  28204: contig of 1780 bp in length
  31215: contig of 3011 bp in length
  33382: contig of 2167 bp in length
  36316: contig of 2934 bp in length
  40441: contig of 4125 bp in length
  44615: contig of 4174 bp in length
  49292: contig of 4677 bp in length
  53647: contig of 4355 bp in length
  59590: contig of 5943 bp in length
  61868: contig of 2278 bp in length
  67152: contig of 5284 bp in length
  71512: contig of 4360 bp in length
  76928: contig of 5416 bp in length
  82257: contig of 5329 bp in length
  89418: contig of 7161 bp in length
  103318: contig of 13900 bp in length
  103319: contig of 9888 bp in length
  113207: contig of 19949 bp in length.
FEATURES
  source
    Location/Qualifiers
      1..133155
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone="RPC198-2017"
BASE COUNT  38122 a 28987 c 27853 g 38141 t 52 others
ORIGIN
Alignment Scores:
  Pred. No.:      5.7e+03      Length:      133155
  Score:          52.00         Matches:    11
  Percent Similarity: 65.22%    Conservative: 4
  Best Local Similarity: 47.83%  Mismatches: 8
  Query Match:     40.31%      Indels:    0
  DB:              2          Gaps:      0
US-09-847-539A-6_COPY_59_86 (1-28) x AC011910 (1-133155)
QY  4  LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db  99660 CTGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAGCGCCGAGTCTCTCTGTGTGGCA 99719
QY  24  AlaAspAsn 26
Db  99720 GCCGACAAT 99728
RESULT 26
AC007416
LOCUS
  Drosophila melanogaster, chromosome 2L, region 28B-28B, BAC clone
  DEFINITION BAC08101, complete sequence.
  ACCESSION AC007416
  VERSION AC007416.12 GI:14327741
  
```

```

/db_xref="taxon:7227"
/chromosome="2"
/map="28E1"
360
/note="push3420 P element insertion site"
join(370..513,695..789,913..1325,1581..13747,3808..4335,
4404..10389,10447..10651,10701..11393,11455..13118,
13175..13657,13715..15650,15706..17302,17366..17584)
/gene="push"
/product="pushover"
370..17584
/gene="push"
join(395..513,695..789,913..1325,1581..13747,3808..4335,
4404..10389,10447..10651,10701..11393,11455..13118,
13175..13657,15706..15715)
/gene="push"
/codon_start=1
/product="pushover"
/protein_id="AAD20449.1"
/db_xref="GI:4426610"
/translation="MSAHSGGTDWNSVVKALINRTGALKNKNEVVNLLKATTCRCHDF
FEENSTOFVTAFAAALADKLMQIKTICOTQICQLHDAVAFIRFIILPRVSVYE
TKWLLGALKMCEGREGCPASASSMFDYNANVVKCKHPHSTTKSIMPSSSSGG
ASNDKSPKSKSDLSVLIQQLIAPLEPKGMTWPLSEEVTDCTEQILAAVNE
YFQONGVDTLDDVCSLPILNRYRSKYMETINGKSLYPLTQVEATAVKSNNHML
TDLTILSOAQAIIEMOPLTPSRIERLSKMGIAALNAVLTISATSVLCMSOASSQKO
TASTSOGSGVGGSGGSKNDHDFEDQACSIVNKALEIYNSIGHMFKTSARIIYON
HLCYGMILLISIGOGANGASGGSSDSASKSKATKSGSEAGTAPTPIARNLF
KVOQFGLNNAIANHSKLLSELIELDKVEACGQSLESTELPEPAQDTLQWSSL
ERIVVLTALHQLFTFLATVAYRKACTLAKRAKOTCEPEISYSDSTYFNDSLS
CSNDEEDSESYLGHWKFTLSPETHDDNANTSTQERAEKSAVLPKLDPHFYLDI
AADIFCDQDQANRHMAYQRYVAGVSDQQLMANIIFKDRDVRNEDQSGNA
PASAAGATAGATKWTQTSIRFSGAAGYIHNLISLLEQSLQSLMLQHLSTSPW
STDTNTPLOVYPTSLVVOILLKTPQEKACLSVWHRLINTLVEGVCSSTASD
SDYEDLNIEHAQLLFLHSLNLMQKKSILLTAGGVIRCAEVCGRGISEDPRVNSOI
MLSRLLLEFLMKHLYNPPELLQVQWNLFSVSSMPDFQKIDLLNKLKNSYC
RHDIENFRSGAGYSGIRPTFYSLVNGDPETSYWAOEFKLDLANNFILCTPKLK
YPLVDALDTLSTSDMSWYKSKDEKASHNLCAIOYCTIAWKLNLGPPSTSHVE
SLKAERSPNLSHMLWSIRLPLASSHYLVVSSLIKQGMTOYAEIWLTHVGDIGADIKY
SLKOTILGEAFNSONGMGTPLRSLDILFSLVAHMQAVANAKGKLPKRESEDA
AGEOSAGTSLTSSNDPELYSSNESIDDKLKDDCKLSSDLQKYQNLVNEMLKLM
NSYQLSVTVRGOMKQLSSSTPEKALNIVPVSDFAPALMLEHAFLKLLPNEDKQ
LIANEWPRCLMNFDAFNGKHPVEYILNVIDAIIIEHTGNSYNTLTLKHLCKSI
LOFELLLELTHANAEVETQQLLISMLDMRTDYLGHSEHCLRETLISGLTQEAQ
LLLEHYMYGYCYRMKLEFAADLRQPAAGPSGAPLQODRAMFNESMLFVAKTWMFKM
LEKVPVQAMROFTQDQSGSLTLLSTFTSLPVSARKMLQFVNLFLQSLQADS
QFOHEDLVECFSELATVDVARKQWGLHLYGPNVSTVSTSEALDTTCMLTSLQOP
SSSSNSAQTPTNATVSMSPSISQDLPMEIEYDCGTAAGGSGSNPTGGAAANTSO
ILSLWQAAPQPPSESSOASDSDSERNGALLSFVKSVDQSKASQIAPPFOALL
QLGQTLTSPQEGCDFADVLIQIMITLADSPARGHVALFNTLLMLELAKQLPDKHL
KHAENVSAQLRYLSLLOQSIGFRGSRNHPWDDDELQTDIDELYDELAEEGEODSLD
DSEDTLNNKLCTFSQTKPEMNQHWYHCTCNMINTVGVCSVCARVCKGHVSYAK
YGNFCDGAKEDSCQALSRRIGSNVSRDAGTGSYLPSSHMSLLACKKRSLLPVGP
VTRKDSLTNERIAVLKLEPYRETLOHQDQWLLVYRCILEYEDLLPSTNENCMLY
STVCHRRATPAALERLHLLQSPQVTDLMFATLGSQEGAFENVMNMSGOGQTIKH
LLSGVVRVAFASLSPHRRQOLAVSHEKVTILQLSALLKOADAQRKLTLQTL
SSAPTACTVLSLAAPNCNEDCLAVGCKECHILTFSSSGSTNEHIVNPPLENGYIK
KAVPLMGQTLTLLAVTSDYKIIDLVADTVSPKYIYLVAVGKIRDCYFVYODGYNML
TFASSGYLTQLODQSLAVHGVFTVNTLLESHQIKDIAHGIGGGGVSYYSHTLQ
LLFYSYSGRSRFPPLTNVSEGVGIYHLDTSKSSKSGKPLQPLQVQTEVTHGGL
VYASMTQNSPFIIMITPERIYLOEIKQAQSAKSRIMDVGRHVAQVEKTLTLLCE
DGSURISAOQPEYTSFMLSQVQPFQNLQYSLTLMKGGSGSTSKSKNSTASGKMTS
KASOQKQPTAGQVPPPIDFEECHNMLADVEFGNDLQIYNKQLKTLFTSTGMF
VASTRNGKFTLVNNDPNVYIGVIRITQDQVQAPSVTILGRTIPTVRARWF
DIPLTRERMLQSKLLKVVFAKAPDEPHVTLLDCELVYKSKELVGVMPDESEDTVPG
SSAPVSSQASSANFEGECNITQDRNANHLLVMDCALHLLGSGVPASMKQAVK
TASALLLTPNPVQTAQRYVLYATLYGTRALYHNKDGVLQFVNLRELSQPKLEKL
ETUREIDPEAFYRLVLRVGTANARPOSIAKICVENNYDIVPTLMVLELHVYTPTL
DEPNVYKRGLCQPEITVHCLVEIMYGFPALDPQVGRMTKYFIDLLKHASVTSLSHA
KEALITLSPKMRKKAIVTTPACSTPTSTSTMOALQAAASAAADIIEAAGVVV
DGSVGGGLPEPNADAEGAAYGVGVGQOQMLNEAFMGGGFPRLGLLPEDADEAIM
DIALALSQQHGGDANALQSQGLANLQGLRQATMAAAVNAANVSLGGSDDEGS
NVATDGSSTLRTSPAEPASGGSGSGVESIGGTSARSSNFGDFHANASPPRGSTKD

```

```

DOEQPSPGVAGSGGVAVLMSSSSEDEANEDDKLSKLDHRLIAVLIESIIQHLGTFD
LNGLOATPLIQVILMLTTDLNGNNEQDQVLDHLLTALVDYVEIGKGAARMETKC
PGNEVRLALLSLFGLMGKTSKGTGTPPHOKDNSSFYASTANVLSKGAIFYA
LEALNTLLVHKNVGLDPIYAGGLASQSAQSGGSGVQLLKPICKGPKPDISIL
IPHNLYKNPDIFESYDGLLTEIIVRLPYQILRLSSAHPDNDYSGFCAMFTTICEYM
MLNNTLLRRQVRKLLMYICGSKFRMYRDSHDAHFYRVKVCNIVSKTGPYIN
ANPMLSYDALVELTEHLTCQEIQSMTGKNQKFCVVHEDALAMLEIACYQLDDGV
SPITIIQLQAAVAVLPPSGSQAPQPSSTSGKLTREDKSDTDAYTSKDPDAQ
GTEVHOIFRYACDALIIRFVRIELLENNITQLRWAHSFMTGLFEHANEROKLLA
IPWNLWPLVPTYGRTAQFVLDLGLYLTSLSTRSITERLPEFVSRAVDVLRQONELLCKH
PNAPITTTLSILQVNGYILESEFCVNCNPEVMANIKLFSVSDSYTITTIYIKL
VQCHTISKLIVRIADLRKMTVNTVYNNRSVOAVVELKNRPAHMKARSVSLSQA
QTEKIDFPLPTACNLMEFADFETVSGSSENIQPCRCASAAFPAYPGVCGNGENY
FOCHKRADNDEKDPFCHSGCFKAKFDFSMYRVCCADVPTESAEADRVTLSLI
HSSLERADNRYROLLTNKQMLLELIQKVAEHRSSDRWDENMASVHSTSOYKNKLIQLL
AQKCVESRSTFEELS KIVQKACRSELVAYDRQOQDQPPVNFQSTTGAENPTNRC
YGCALASTEQCLTLRLWAYNTHCRVCLYSOGLVSELAENLRGTPLIQEVRNLLY
VLTKDNAPACHMLLQVLTTRVKNALMGSIPLISLEAAVHOEMTLLLEVLLGDDICWEY
KLKVFELFISNCRTL*
3135
/variation
/gene="push"
/replace="t"
3600
/variation
/gene="push"
/replace="t"
17558..17563
/polyA_signal
/gene="push"
17584
/polyA_site
/gene="push"
4795 a 4143 g 4406 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 674 Length: 17595
Score: 52.00 Matches: 11
Percent Similarity: 65.22% Conservative: 4
Best Local Similarity: 47.83% Mismatches: 8
Query Match: 40.31% Indels: 0
DB: 3 Gaps: 0
US-09-847-539a-6_COPY_59_86 (1-28) x AF096896 (1-17595)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
|||||
Db 7027 CTGGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCAAGCGCCGCGTCTCTCTGTGGCA 6968
|||||
Qy 24 AlaAspAsn 26
|||||
Db 6967 GCCGACAAT 6959
|||||
RESULT 24
AC018007
LOCUS
DEFINITION
Drosophila melanogaster, 36993 bp DNA linear HTG 09-DEC-1999
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
places.
AC018007
AC018007.1 GI:6553183
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 36993)
ADAMS,M. and Venter,J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212995 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and

```

MLLSRLLEFLLEYLMKLYNAPPELLDQVRWNLFPVSMSMDPTQKITDILLNCRTKLNSVC  
RHDIEENFRKSEAGYESSIRPTPEYSILVMDGPETSYWAQEFKLDGLAWNTEILCPDCLK  
YPLLYDALDILSITDSMSYSEKEDREASHHNLCAQCYETIAWKNLGLPSTDSHVE  
SLKAERPNHLSUMWISIRUPLASSHYLVSSLLIKQGYTQYIAETLTHWVGIDIGADIKY  
SKQOTLIGVEAFNSONGNGTPRLSDLIIFDSLVAHQAVAWANKKEGLKRWELKESEDA  
AGQSAGTSILYSSNDPELYSSNESIDDOKKQDDCKLSSDLOKYNQLVNLWELVKIM  
NSYQLLSLVIRGOMLKQLSSSTPEKALNLIPIVSDKPAIMLELHAFLKLLPNEDKO  
LIANEWPKLVNDRFAFNGKHQHPVEPIILNVIDAHLIELTRGNSYSLTHPLKCLKSI  
LQJPELLPHRTANAETVETQKOLLSSMLDMRTDYLQHSERCLREILSGLTEQAK  
LLELTHMGVGYRMLKEFADLRQPAAGSGGAPIDQDRAMFNESMLFAVLMTKFM  
LEKPAVAMQRFQKQDQSGSTLTLISFTGSLPVSYARKMLQFNRFLQLSLOADS  
QFQREDLFCFSELAETVARLKWGLGHIYIGPNVSTDVSTSEALDTCRLMSTILQ  
SSSSNAOTPTNMTVTSAMPSISDOLDHMEIYDCGTAAGGSGNPGCTGAANAATSO  
ILSLWQAQDPDEESSQACDSDSERNGALLSPVSLVKQDSKASQIAPPLFOALL  
KQGTLSIPQBCPADVQLQIMITLADSPARGHVALFNTLILWLELAKLQDOKRHL  
KHAEYSLAYRLELSDQLSIFGRSROHNPMDDELQTDIDELYDELAEEGQSDLLD  
DSDDELNNKLCFTSQTOKEFMQHWYHCHTCNMINTGVCSVCYARCHKHGDVSTAK  
YGNFDCDCAKDCGSAALSRRTGSNEVRDSAGISGLVPSHMSLLAGKRSLSLPVQOP  
VLTRKDSLNERIATVLTKLLEPEVRETLQHDOWLLVVRCLILEYFDLILLPSINENCLY  
SIVGCHKRATALERLHLLQSQVTDQLMFATLQSQEGAFENVRNYSDDQOTLKH  
LLSGVVRVAFCLSPHGRQOLAVSHEKGVITLQLSALKLQADASKKULITQL  
SSAPIACTVLSLAENPCNEDCLAVGLKECHILTFSSGSTEHIIVNPQLENGNYIK  
KAVLPGSOTLAVVTSYKIYDLAVTSPKYYIYLVAVGIRDCDTFFVQDGNMYL  
TFASSGYITQDQOSLAVHGDYVYNTLSELSASKSASKAGIPIQGLVQVTEVTHGL  
LLFYSYSGRSFSPSTNVSEGVKGIYHLDTNSASKSASKAGIPIQGLVQVTEVTHGL  
VASMOTSNPILMLTPEPIYLOEIKAKSASRIIMDVYGIHRAVAGVETLILLICE  
DGLRIFSPAGTFTWLSPOVQFGNLOLSSYLSLMAKGGSGSTSKSNTASGKMTS  
RKASQOQKQAPAGQVFPIDFHECNMLADVEFGNDLLOIYNKOKLKLRTFSTGMF  
VASTRNGTELEVINDPNVIVIRVIGTQDQVAPQSVITLIGRTIYVPPRRARF  
DIPLTREEMQLSKDLKLVAKAPDEPHVTLLDCEIYVGSKELGWPDSEEDVTPG  
TSAPAVSSQASANFGGNCITOLDMANHLLVENDCALHLLSGCVAPMSQKQAVK  
TASALLPTPTPOQARYVATLITGTRALYHNYKQVLLQFVNRLOESQKPLEKL  
ETUREIDPEAFYRLVLMVRIANRQPSLAKICVENNYDIVPTMLVLELHVPTPL  
DEPNVIVKGLQPEITVHCLVEIMYGFALDPQGVGMTKYFIDLLKHSISISA  
KEALLILLSKPKRRVAVITPPACSTPTSTMQALQAASASDSDIDEEAGVYV  
DGSVGGGLPEPNADAGAAVGGVGOOQMLNLEAFMGGGPRLGLLPDADDEATM  
DIALLSLOQHQDANALQLOGLANLQIGIROCATAMAAVANAANAVSLGSDDEGS  
NVATDSTLTSPEAPGSGSGSGSGSVESIGTSARSNFSGDHANASPROGSTKD  
DQOQPSGVAGSGVAVLSANSSSEDEANEDOKLSKLDHRLITVLESTIQLHGTED  
LCNGLAQIPIILMTTDLNNGERDQVLLHLLTALVGLGCKNKGAAARATKTC  
PGNEVRLALLSVGLMGKTSKQGTGTSPPHQFKDNSFVASTTANVLSKSGAFYA  
LEANTLRLVHKWGLVADYAAAGGLASQASAGSAGSGVQLLPIKHGPKPDISTL  
IPHNYLKNYDPIFESYDGLTEIIVRLPYOILRLSSAHPDNYDSGCEAMETFLCEVM  
MLNLTLLRQVRLKLLMYICGSEKERMVDRGSLDAHFVVRVRCVNYSSKTCAPYN  
ANPMLSYDALVELTEHLRTCQISQMRTNWOKFCVYHEDALAMLEIACIQLDDGV  
SPIIILLOLAQNALPPPSKQAPQPSSTSSXGLRTDREKSEDTDAYTSKFPDQ  
CGTFVIOFRYACDALIIFRIFLLENNITQLRQAHSMTGLFEHANRGRGLIN  
IPWNLVPTYCGRTAQFVLLGLYTLSTRITERLPEVSRADVLRQOONELCKH  
NPAPIVTTLESILQVNGYXLESPECLVCPNPEVPMANIKLPSVKSDSKYTTTMYKL  
VQCHTISKLVIRLADLKRTKWRTIIVYNNRSVQAVVELKNRPAHKKARSVLSQA  
QTEKIDFPLPITACNLMIEFAFFETVSGSSENLOCPRCSAVPAYPGVCGNGENY  
FOCHKRAINDKDPFLCHSCGFCYAKFDFSMYARVCCADVPIDESAEDRVKTSLI  
HSSLERADRYRLLNKNQLELLIQKVAEHRSDRMVEDNMAVSHSTSOVNKIOLL  
AOKYCVESRTSFEELSKIYOKVACRSELVAYDROQDOPPVNPGSTGTAENPTNRC  
YGCALASTECLTLIPAMVNYDCRVCLISQGLVSELAENLRGTPLOIEEVRNLLY  
VLTKDNACAEHLQLVTRVKNALMSIPLISLEAAVHQMELTLEVLVQGDIDCKMY  
KLKVIPELFSNCLPRGPVAVLHPCLRIMQNLICPVLPSPKPNQVAVTDLCSMKI  
FEGNTVYSEKYNDRNHEYAASKRPSNNQAKLNKAKDQNAAGSGDAPKRSRE  
VRVAFYSEKYGKWRERVDLQKRVIPLVNAKWIOPLFNANSRQVLOACLSLSL  
SRTNERRQOALNMTSFLKHVGEAGEASAEYLMYKNMATEQPMLOYLYTKMLGVLSLS  
OLLAIEISKVHMEEYISLSDISLGYALRQYVELLWLLLECPNIRTYTKRMGPVLE  
SYIALSLSVQVRLTDDAQEKLEMDLETGTETEETRAFNEILIDTVEKTRMNDIK  
TPVVFERYSIITHEPEHDESEFTMLEKDQDEDFLQGRMLGNPVPSSBMGLDMLR  
DVNKKITCELLALDEDNGMELLNKKIISLDLPVKYDGYIKVWLAEGDGRAMRIV  
YMRGLLOGATEFEETVNNKSOQVDETOELRYMANVLADCNKLOSENDESIOAVTE  
RNRELLOVLLKFLICVKVRCOEVLCOPEIGAINITLLKVLQMCLOSENSIOAVTE  
QLEIEMETILSKAASITLDSFLQSLTFGQPEYVSALISCTDCPNVRNPNVRLHLIR  
VLAALYGNVKNALICEHFKDTLNRPNRPNERTPEEFKLELFCVLTNOLEHNCIGG  
TKLDYIVSUGVAVAIKSDIPIIHLREQVSSDEHVSLEAENLEALSTDSATAARQVQVDFPT  
RAEKKRLMATREKQIDALGMRTNEKQVTAKSIILOKIEKLRDETGLCTFCREGYA  
COPDKVLGIYTFKRCNVEPEFLKSRKTIGYTTVTHFNVVHVDCHTSAIRLTRGRDWM  
ERASLQNAWTRCNGLPLMGPAVGEAFAFSACMTRHSSYMQESTQRCDISTYSVHDLK

LLLVRAWERSPHDAGGGGQSQNMHPVPYLLFVSVVLLLSRSSAARDSTKLITYLQA  
PPEKWLCECYEVDGPLFMATISLSLHSERLNKKHLAHLKRLMIAVAOQHIVSPAVLC  
KALLADPADROYKDYTYKPELMMALVDLIDNLFKTYSTPKEDWPISLFDYLRKND  
EALLKSTDSLQTLTTEFLPCTSFVEFCOVAGLULHLEHPDNFTEELAAALPSTSSN  
\*  
variation 2207 /gene="push" 16135  
/replace="t" 11  
2672 /gene="push" 4  
/replace="t" 8  
5411 .5599 /gene="push" 0  
/note="putative" 3  
/bound\_moiety="zinc" 0  
11402 .11542 /gene="push" 0  
/note="putative" 0  
/bound\_moiety="zinc" 0  
15995 .16135 /gene="push" 0  
16130 /gene="push" 0  
/note="putative" 0  
polyA\_site 16130 /gene="push" 0  
BASE COUNT 4288 a 3943 c 4035 g 3869 t  
ORIGIN  
Alignment Scores: 615 Length: 16135  
Pred. No.: 52.00 Matches: 11  
Percent: 65.22% Conservative: 4  
Best Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 3 Gaps: 0  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x AF096897 (1-16135)  
Qy 4 LeuGluAlaLeuAlaAspGluThrAspAlaLeuGlnSerGluAlaAlaValVallys 23  
||||| ||||| ||| ::::: ||||| ::::: |||||  
Db 5971 CTGAAGCTGTCTTCACGACGAGTGAGTCTCTCCAAGCCGCGAGTGTCTCTTGTGTGCA 5912  
Qy 24 AlaAspAsn 26  
||||| |||||  
Db 5911 GCCGCAAT 5903  
RESULT 23  
AF096896/c 17595 bp DNA linear INV 15-OCT-2001  
LOCUS Drosophila melanogaster pushover (push) gene, complete cds.  
DEFINITION Drosophila melanogaster pushover (push) gene, complete cds.  
ACCESSION AF096896  
VERSION AF096896.1 GI:4426608  
KEYWORDS fruit fly.  
SOURCE  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephyridiidae; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 17595)  
AUTHORS Yager, J., Richards, S., Hekmat-Scafe, D.S., Hurd, D.D., Sundaresan, V.,  
Caprette, D.R., Saxton, W.M., Carlson, J.R. and Stern, M.  
TITLE Control of Drosophila perineurial glial growth by interacting  
neurotransmitter-mediated signaling pathways  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (18), 10445-10450 (2001)  
PUBMED 11517334  
REFERENCE 2 (bases 1 to 17595)  
AUTHORS Richards, S., Hekmat-Scafe, D.S., Hurd, D.D., Caprette, D.R.,  
Saxton, W.M., Carlson, J.R. and Stern, M.J.  
TITLE Direct Submission  
JOURNAL Submitted (06-OCT-1998) Biochemistry and Cell Biology, Mail Stop  
140, William Marsh Rice University, 6100 South Main, Houston, TX  
77005, USA  
FEATURES  
source Location/Qualifiers  
1..17595 /organism="Drosophila melanogaster"

MLFAVLTKMFKLEKPVAVQAMROFFKQDRSGSLTLLLSFTGTSPLVSVARMLQPV  
NRFLQSLADSOQHEDLEVECFSELATVDVARLKWGLHIIYGPVNVSTVSEALD  
TTCMLSTIIQOLPSSSSNAQPTTNATVAMPISIDQLDPMEIYDCGTAAGSEGN  
PGTGAANTQILSLMQAAQPNPSESSQACHSDSERGALLSVFKSLVKDQSA  
SQIAPPAQLOLQLOGLTIGPPOEGCDFADVLOMITLADASPARGHVAFLNTLLWL  
ELAKLOLPDKHLHAENVSAQRLVLSLQISIGRSGROHNPWDDLOVDDELDE  
LAERGEDSLDSDSDTLLNKCTFSQTKQEPNQRWYHCHFCNMLNTVGVSCVAR  
VCHGHDVSTAKYVPCDCGAKEDSCQALSRKNGSNEVDGAGISYLPUSHMLIA  
KKRSLPQVPLTNRSLTNERIAVTLKLEPYRETLOQDWLWQCEILEYFOL  
LLPSINENCQVLYTVCGRKATRALERHLLEQSFQVTDQMEATLWQCEILEYFOL  
NYSQDQOTIKHLLSSGVWRVAVPCLSSPHGRQOOLAVSHEKGVTLIOLSALLKOA  
DASKRLITLQLSAPLACVVISLAANPCNEDCLAVGLKECHILTFSSSGSNEHIV  
VNOLENGTIKSAWLPQSGTISLAAVVTSDYKIYDIAVDTSPKTYLLVAVGKIRDC  
TFVYQDNYNMLTFASGYIYQOOLDOSLAVHGFYVNTLLESHOIKIDAGHIGG  
GGVSIYSHFTLQLLFYISYGRSFPSTNVSEGVKGIYHLDTNASAKSKGPIQL  
VOMTEVGHPLVYVAMOTNNPILMITPIYIYLOBIKAQSAKSRIMDVVGRHAVA  
GVEKTLILLCEGSLRIESAOPETSTFWLSPQVOPFGNOLYSTLMKGGSGSTSK  
SKSNTASKMTSKASQOQKOPTAGGQVPEPIDEFHCNMLADUEFGCNLLQIYNKO  
KLKTRLFSTGHFVASTNSGFTLEVINDPNVIVGIRVIGTODVORAQPSVITLGR  
TIPFVRRARWFDIPLFREEMLOSLKLLVYFAKAPDEPHVTLDDCEVYKSKELVG  
WPDESEDVTPGSSAPAVSSQASSANFEGFCITQLDRMANHLLLEVMDCALHLLGS  
GVPSMOKAVKATASALLPTPNVOTQARYLATITGTRALYHNYKDGVLLOFVNR  
ELOSMPKLEKLETLREIDPEAFVRLVLMVGTIANARPOSIAKICVENNDIVPTLMG  
IVLELHVTPIDEPNIVKRGLOCPETIVHCLVEIMYGFALADPGQVGMKTFIDIL  
LKHDASVISHSAKALLILSPRMKRRVAIVTPACSTPTPTSTWQALQAAASSAA  
SDIITEAGVYVDSVSGGGLPEPNADAEGAAGVGVGQOQMLNLEAFMGGPPRL  
GLPEDADEALMDIAIALSIQOHHGDANALQSPOGLANLQGTQATMAAANVAAAN  
VSLGSDDDGSENVATDGTSLRTSPAEPAGSGSGSGVESIGGTSAKSNFGDHA  
NAPPSROGSKDQOEPGSPGAGVAVLSAMSSSEDNEANEDDKLKLHLDIRAV  
LESIIHLGTFDLGNGLOAIPLOVILMLTDLNGNRRQOVLHLLTALVDEIG  
KVASARNEKTCNEVRLALLSLFVLMGKTSKQGTGTSPPHOFKDNDSFVASTTA  
NRUSKGAFCVPEALANTLLVHNKVLGDPYAGAGLASQASAGSGAGVQOLLKP  
IKHGKPTDISILPHNTLKNYPDIFESYDGLLTIIVRLPYQILRLSSAHPNVDYSG  
CEAMFTPLCEYMLNLTLLRQVRKLIMICGSEKFRMYRQDGLSHDAHFVVRKYC  
NIVSSKGTAPVNPMLSYDALVELTLEHLCOEISQMRGTGNWKCVCVHEDALML  
MEIACYOLDGVSPIILLOLQAAVCLNPPSGSKQAOQPSTSSAKSLKTRDEKSD  
TDAYSFDPACQCTVHQIFRYACDALIIRIFRIFLENNITOLRQWASFTMGLE  
HANERQELKHPNAPVPTTLEIYQVRYTAQVLDLGLTILSTRILPEFVSRAVD  
VLRQNELKHPNAPVPTTLEIYQVRYTAQVLDLGLTILSTRILPEFVSRAVD  
SKYTTMTYLVQCHTISKLIVRIADKRTKVRTINVYNNRVOAVLKNRPL  
WHKARSVLSAQLEKIDPLPITACNLMEFADFFETVSGSENQOCPCRSAAVA  
YPGVCGNENVOFCHKRAINDEKDFLCHSGCFCKYAKDFSMYARVCCAVDPIE  
SAEDRVNLIHSSLERADNRNQLTNKQMLLEKLOKVAEHRSDRMVMDNNAVSH  
STOVNKIILQALQYCEVSKTSFEELSKIVQKACRSELVAYDROQOQPNVNGS  
TTGAENPTNRYCALASTEQCLTLRAMAYNDRCVCLYSQGLVSELAEHNRLRT  
PLIOEVRNLIIVLTNDNAPACMLLQVTRVKNALMGSLPLISLEAAHOENLTLE  
VLQODDTCWEYKLVIFELFISNCRLPGRPVAVLHCLRIMONLIPCVLPGSKPNO  
KVATDLCMKMFGNTVDYRAWLNDRNHEYAAWSKRMPSNQAQKLNKAKDNVAAS  
GGSADPKSREVRVAFLESEKIGRWREVRVLDKORVIKPLVFNQAQLOPLFNANSRP  
GRQACSLLSLSTRERROOALNMLTSLFKHVGGEAGEASAEYIMLYKNMATEQPLQ  
YLVKGLVLSQISLQALATEISKVHRMEYSLSLSDLSGLAYROYVELLWLLCEPNIR  
TYKTRMLGPVLESYLAIRLVQRTLIDDAQEKLEMLEDMTSGTEETRAFMETIL  
DTVEKTRMDIKTVPVFERLYSIIHPDEHSEFYNTLEKDPQOEDFLOGRMLGNPY  
PSEMGGLPMDRVKMKICUCLEIALLLEDNGNELLYNNKIISLDLPVRYKKVWL  
AEGGDRAIRVPMRGLGLDTEEFVLRNKSQEQVTEQLYRMANVLADCNGLRV  
MLERIGSLRISNRRELIOVLLKFLICVKVRKQCEVLCQPEIGAINLLKLMQMLQ  
SENDSIQASVLELLETMETILSKAASDTLDSFLQSLTFEGGVEYVALISCTCPNV  
RNPSPVLRHIVLAAALVYGVNEKMALICEHFKDITLNFNRDNERTPEEFKELFCV  
LTNOIEHNCIGLKDYIVSLGIGERSLAYITHAPCVKPTLRLTSDDELKEPSTRPS  
LKYTLRFLTGLSNHEATOVAKIDIIPIIHLRQVSDHVGSLAENLLEALSDSA  
TAARVQVRETRAEKRLAMATREKQDALGCMRYNEKQVATAGSILQKTEKURDET  
GLTCTCREGFAVQVGLVITFTKCNVEEFELSKRTIGYTTVTTHVNVHVDCT  
SAILTRGDEWERSALONTRCGLPLWPAVGAAFSACFTSRHSSYVMOESTQRC  
DISYTSVHDLIKLLLVFAWERSFDGAGGQPMHFPVYLLFYSLLLSRSAA  
RDSYTLTLOQAPSEKWLCEGYEVDGQFLFMATLSLSHSELNKKHLAKHLRMIAV  
AQRHVSAPVLCALLAPADQVKTVPKPLMMAVLYDLYDNLPKTVSTPREEDW  
PIASFDLRKNDEALLSTUSILOTLTEEFPLCTSFVFCVDAGLLHLHIEHPNFIEE  
ILALPSTSSSN\*

BASE COUNT 3257 a 3028 c 3148 g 2940 t

ORIGIN

Alignment Scores:

Pred. No.:

465

Length:

12373

Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 3 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x DME17920 (1-12373)

Qy 4 LeuGLuAlaLeuAlaAspGlnThraspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
||||||| ||| :||| ||||| :||| |||||  
Db 2329 CTGGAAGCTCTGTTCACAGGTCGAGTCTCTCCAAAGCCGCGAGTCTCTCTTGTGGCA 2270  
||| ||| :||| ||||| :||| |||||

Qy 24 AlaAspAsn 26  
||| |||

Db 2269 GCCGACAT 2261  
||| |||

RESULT 22  
AF096897/c AF096897 16135 bp mRNA linear INV 15-OCT-2001  
LOCUS Drosophila melanogaster pushover (push) mRNA, complete cds.  
DEFINITION Drosophila melanogaster pushover (push) mRNA, complete cds.  
ACCESSION AF096897  
VERSION AF096897.1 GI:426609  
KEYWORDS fruit fly.  
SOURCE Drosophila melanogaster  
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 16135)  
AUTHORS Yager,J., Richards,S., Hekmat-Scafe,D.S., Hurd,D.D., Sundaresan,V.,  
Caprette,D.R., Saxton,W.M., Carlson,J.R. and Stern,M.  
TITLE Control of Drosophila perineurial glial growth by interacting  
neurotransmitter-mediated signaling pathways  
Proc. Natl. Acad. Sci. U.S.A. 98 (18), 10445-10450 (2001)  
JOURNAL 11517334  
PUBMED 11517334  
AUTHORS 2 (bases 1 to 16135)  
REFERENCE Richards,S., Hekmat-Scafe,D.S., Hurd,D.D., Caprette,D.R.,  
Saxton,W.M., Carlson,J.R. and Stern,M.J.  
TITLE Direct Submission  
JOURNAL Submitted (06-OCT-1998) Biochemistry and Cell Biology, Mail stop  
140, William Marsh Rice University, 6100 South Main, Houston, TX  
77005, USA

FEATURES  
source  
1..16135  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2"  
/map="28E1"  
<1..16135  
/gene="push"  
<1..25  
/gene="push"  
26..15994  
/gene="push"  
/codon\_start=1  
/product="pushover"  
/protein\_id="AAD20450.1"  
/db\_xref="GI:426611"  
/translation="MSAHSGGTGDNVSVKALILNRTGALNKNVNVLLKATTCRCHDF  
FEESNTOFTAFALAAADKLMQIKICQICQLDHATAVLIRFIYLRPRVSYVE  
TKWLGLMLCEGREGCPASSSMNFDNVANVVKCKHPSTTKSIMPSSSSSGS  
ASNDKEPKSIKRSRDLSSVLIQLIAPLEPGKMTWVPLSEETDCTQILAAVLE  
YFQNGQVTLDDVCVSLPILNRYSKYMETINGKSLYLPQTVEATAVKSMNHL  
TDLTLSQAOLAIEMOPLPSRIERLSCMGTAALYNAVLTISATSVLGMSSQSKQ  
TASTSQSGVGGSGGOSKDDHDFEQACISVANKALEIYNIGHMFKTSARIVYON  
HLVCGSWLLISGICGACAGSSGSSSDASAKSKATKSGSEAGTAPTTIAPVNL  
KVOQGFELNANIANHSTKLUSLIEDLKVCAACQSLSELPPEAQFDILQNYSSL  
EIRIVPLNLTALHQLPTTATVYAKKATLKRASKDRTECEPISYSDTFTYNDLS  
CSDNSEEDSESYLGHWFKETLSPETHDDNANTQERAEOKSALVPLKDPHEYLDL  
AADIFCFDQFLANRHAYMORYKAGVSDQOMLMANIKDFDRVMRNETDQSGNA  
PAASAGAGTSAGTKWQTSMLRFSGAAGRYIHNLISTSVLSEOLSNLLOHLSIPW  
STDTNTPLOWYPTLSVLVQIILLKLPQKEACLSVWHLINLTIVEGVCSSTASD  
SDYEDLNIEHAQLLLFLFHSUNLMOKKSIILLTLAGGVIRCAEVCGRGISEDPRVNSQI







misc\_feature /note="assembly\_name:Contig51"  
 120563..130242  
 misc\_feature /note="assembly\_name:Contig52"  
 130343..140588  
 misc\_feature /note="assembly\_name:Contig53"  
 140689..149261  
 misc\_feature /note="assembly\_name:Contig54"  
 149362..162349

Alignment Scores:  
 Pred. No.: 5.23e+03 Length: 173772  
 Score: 53.00 Matches: 11  
 Percent Similarity: 77.78% Conservative: 3  
 Best Local Similarity: 61.11% Mismatches: 4  
 Query Watch: 41.09% Indels: 0  
 DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC087862 (1-173772)

Oy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21

Db 173045 CTGGAGGCTCTGCTGGACAGACAGACGGGCGACAGAGCAAGTAGCCTTG 172992

#### RESULT 20

AE004659/c  
 LOCUS AE004659 12411 bp DNA linear BCT 30-AUG-2000  
 DEFINITION Pseudomonas aeruginosa PA01, section 220 of 529 of the complete genome.

ACCESSION AE004659

VERSION AE004659.1

KEYWORDS GI:9948361

SOURCE Pseudomonas aeruginosa.

ORGANISM Pseudomonas aeruginosa

Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

Pseudomonas.

1 (bases 1 to 12411)

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K., Wu, Z., and Paulsen, I.T.

Complete genome sequence of Pseudomonas aeruginosa PA01, an

opportunistic pathogen

Nature 406 (6799), 959-964 (2000)

20437337

2 (bases 1 to 12411)

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K., S., Wu, Z., Paulsen, I.T., Reizer, J., Sailer, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.

Direct Submission

Submitted (16-MAY-2000)

Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

Location/Qualifiers

#### FEATURES

source 1..12411

/organism="Pseudomonas aeruginosa"

/strain="PA01"

/db\_xref="taxon:287"

complement(353..1117)

/gene="PA2327"

complement(353..1117)

/gene="PA2327"

/codon\_start=1

/transl\_table=11

/product="probable permealase of ABC transporter"

/protein\_id="AAG05715.1"

/db\_xref="GI:9948362"

/translation="MRQTLRGALGVAGLLTLFALWLVGALFGAAGDLSARSPQAT

LLSLGELLERRELYEHVLVSLKRVLLGLFALLVGVPLGLVGASRNLEAATTPAQF

LRMISPLSWMPIAVMLMGVDQPIYFLLAFAALWPILLNTAAGVRQLDPRWLQLSRL  
 AATPWETLRVILPGLGLVLTGTVRGAIGLWLVPCPEMLGVSAGLYILDTDRLL  
 AYSELMAVLILGLGLYLLDAARGLYRRWAHGAIG"

gene

complement(1114..2313)

/gene="PA2328"

complement(1114..2313)

/gene="PA2328"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAG05716.1"

/db\_xref="GI:9948363"

/translation="MCDUPTHSRRDILKLAALLSAGALPLSLSSLQARAABPDAPV  
 RIGYLPYTDATPLLVHANGLFAEGIQAEPLVLRSWAQVIEAFISGVNVVHLLSP  
 MTVMARYGSKVPKAVVAVNHVGGSLTVAPEDVRLQGLGKSVAIPIWYSIHNVVLOQ  
 LLRLDGLRAVSRAVGAALAEVNLVLPSPDPPALASKRIGHYIVAEFNRALAEAL  
 KVGVRQFTGDMVRNHACCVVHEHDLRRPQWOKVNAIVKAQLWTEHREHRAEAQ  
 LLSKAGANRYTPHAPVGLRVLAGPAEEQQAYLASGAIRHADWQERIDFQPPYPSY  
 TELVRRKDLTLEGNNFLASLDPROVASDLVDREVRQALAEVGGVLAQGLAEGFQ  
 RSEISIP"

complement(2279..3115)

/gene="PA2329"

complement(2279..3115)

/gene="PA2329"

/codon\_start=1

/transl\_table=11

/product="probable ATP-binding component of ABC

transporter"

/protein\_id="AAG05717.1"

/db\_xref="GI:9948364"

/translation="MSEALLDARGIALGYPRGAGQAVLGFELRLGAGEVVSILGPS  
 GVGKSLRLVLAGLQEPAGSVRLGPHGVAFAVDFSLPLMLNLEKNVAFG  
 LDFARQPHLDAATQRAVDQAIAAVGLEHARARYPAOLSGGMAORTALARCLARQPV  
 LLIDPEFGALDEVTRADMQLLRAIHERTGAVVLTIDIDEALLSERILLGLGSPA  
 RTLGWEIDLPQRAELVEELGALRIELKTLRASRTHAHPAQPEASHVPGRPDPF  
 PT"

complement(3112..4179)

/gene="PA2330"

complement(3112..4179)

/gene="PA2330"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAG05718.1"

/db\_xref="GI:9948365"

/translation="MLDPTLSDWLDROADLDRGDADPAALLPLAAGVGLGVAVER  
 LGAGGTLADAVERYAAVARSLSLTAAFVWGQRAFIYLLQSPNQELRQLLGLLAG  
 RLKATGLSNAMKFLSGTEALQVVAHAEAGWRUDGLRPWYNLRSGDFVYAAAEIHG  
 EGGKPFVLAIPGGLQSRADLRLGLQCSNTAALDRAVAGRDMLLHDDARQFLP  
 RVRFPLGLQCGMAIGLARALDEVORHLAGSRSLDGLGELAAQRETLDGHVAAHLACL  
 ASGEFASQPARLFRIRITGLAEAAANAVOLELQASGGKAYLSEHGAGFARRWRESAFVP  
 IVTPSLVQLRAELQROAEAKA"

complement(4229..4789)

/gene="PA2331"

complement(4229..4789)

/gene="PA2331"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAG05719.1"

/db\_xref="GI:9948366"

/translation="MSRPVPLTENAPEARPFLQTLNNSGYIPNLGLVLANAAPAL  
 EYTLTVSGLNAKASGLPEREVQLVAATTHGCDQFCVAGHTAVARNKARLPPEVIEAL  
 RARGELPDARYEALADFTRAVIASRGAVTDEQFEAFRAAGFDQAQALEVILGVSLATL  
 CNFANVTARTPLNAELQYRWTFNQA"

4894..5799

/gene="PA2332"

complement(4894..5799)

/gene="PA2332"

/codon\_start=1

/transl\_table=11

/product="probable transcriptional regulator"

/protein\_id="AAG05720.1"



```
repeat_region /rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
misc_feature 10676..10831
misc_feature 10829..11215
misc_feature /note="similar to EST AW070186 (NID:g6025184) xa09a03.x1"
misc_feature 10829..11250
misc_feature /note="similar to EST AI806840 (NID:g59393406) wf15h08.x1"
repeat_region 11262..11805
repeat_region /rpt_family="ERV1"
misc_feature 12808..13247
misc_feature /note="similar to EST AI908793 (NID:g6499473)"
misc_feature 12809..12925
misc_feature /note="similar to EST AI908824 (NID:g6499504)"
repeat_region 12952..13235
repeat_region /rpt_family="ERV"
repeat_region 13237..13580
repeat_region /rpt_family="MaLR"
repeat_region 14241..14593
repeat_region /rpt_family="MaLR"
repeat_region 16242..16534
repeat_region /rpt_family="Alu"
misc_feature 17189..17450
misc_feature /note="similar to EST D62720 (NID:g966494)"
repeat_region 18011..18146
repeat_region /rpt_family="L1"
repeat_region 18705..18738
repeat_region /rpt_family="L1"
repeat_region 18739..20933
repeat_region /rpt_family="L1"
repeat_region 20934..21246
repeat_region /rpt_family="L1"
repeat_region 21247..21557
repeat_region /rpt_family="Alu"
repeat_region 21740..21881
repeat_region /rpt_family="Alu"
repeat_region 21983..22295
repeat_region /rpt_family="ERV1"
repeat_region 23119..23239
repeat_region /rpt_family="MIR"
repeat_region 23639..23934
repeat_region /rpt_family="L2"
repeat_region 24194..24335
repeat_region /rpt_family="MaLR"
repeat_region 24452..24562
repeat_region /rpt_family="L1"
repeat_region 24961..25528
repeat_region /rpt_family="ERVK"
repeat_region 25530..27490
repeat_region /rpt_family="ERVK"
repeat_region 27485..27807
repeat_region /rpt_family="ERV1"
repeat_region 27808..27914
repeat_region /rpt_family="ERV1"
repeat_region 27922..28018
repeat_region /rpt_family="ERV1"
repeat_region 27997..29043
repeat_region /rpt_family="ERV1"
repeat_region 31133..31165
repeat_region /rpt_family="ERVK"
repeat_region 31172..31597
repeat_region /rpt_family="ERVK"
repeat_region 31598..32237
repeat_region /rpt_family="ERV1"
repeat_region 34473..35285
repeat_region /rpt_family="L1"
repeat_region 35883..36498
repeat_region /rpt_family="ERV1"
repeat_region 38463..38929
repeat_region /rpt_family="MER1_type"
repeat_region 39354..39609
repeat_region /rpt_family="L1"
```

```
repeat_region 42594..42877
repeat_region /rpt_family="Alu"
repeat_region 43519..43817
repeat_region /rpt_family="Alu"
misc_feature 45998..46226
misc_feature /note="similar to EST AA909108 (NID:g3048513) om32d11.s1"
repeat_region 47950..48034
repeat_region /rpt_family="L1"
repeat_region 48869..48985
repeat_region /rpt_family="Alu"

Alignment Scores:
Pred. No.: 3.46e+03 Length: 117309
Score: 53.00 Matches: 11
Percent Similarity: 77.78% Conservative: 3
Best Local Similarity: 61.11% Mismatches: 4
Query Match: 41.09% Indels: 0
DB: 9 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AC007274 (1-117309)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21
|||||  |||||  |||:::|||||  |||||:::||||  |||:::
Db 96844 CTGGAGGCTCTGCTGGACAGGACAGCGGGCACAGACAGGAGTAGCCTTG 96897

RESULT 19
AC087862/c
LOCUS AC087862 173772 bp DNA linear HTG 19-FEB-2001
DEFINITION Homo sapiens chromosome RPCI-11 clone RP11-724H13, WORKING DRAFT
SEQUENCE, 37 unordered pieces.
ACCESSION AC087862.3 GI:12963066
VERSION AC087862.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 173772)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 173772)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Feb 18, 2001 this sequence version replaced gi:12745114.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0724H13
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 150537 bases at least Q40
Consensus quality: 158418 bases at least Q30
Consensus quality: 162327 bases at least Q20
Insert size: 191000; agarose-fp
Insert size: 170172; sum-of-contigs
Quality coverage: 2.88 in Q20 bases; agarose-fp
Quality coverage: 3.33 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
```

```

/note="3 base direct repeat bordering transposon Tn5714.
Note difference - TCA in Streptomyces coelicolor plasmid
SCP2, TGA in Streptomyces coelicolor plasmid SCP1."
complement(7784..7794)
/note="Inverted and repeated imperfectly at
complement(11285..11295)"
complement(7784..11295)
/note="Similar to Tn5714 from SCP1"

Alignment Scores:
Pred. No.:      857      Length:      31317
Score:          53.00    Matches:      13
Percent Similarity: 65.38%  Conservatives: 4
Best Local Similarity: 50.00% Mismatches: 9
Query Match:      41.09%  Indels:      0
DB:              1      Gaps:      0

US-09-847-539A-6_COPY_59_86 (1-28) x SCSCP2 (1-31317)
QY 3 AlaleuGluAlaLeuAlaAspGInThrAspAlaLeuGlnSerGluGluAlaAlaVal 22
Db 23713 GCGAGCGCGCGCGGACCGAGCGGCGCGTCCAGTCGGAGCGCGCGCGCGGCC 23654
QY 23 LysAlaAspAsnAlaAla 28
Db 23653 GAAGCGGACCGCGCGGCC 23636

RESULT 18
AC007274 LOCUS
DEFINITION Homo sapiens BAC clone RP11-105L10 from v, complete sequence.
AC007274 ACCESSION
VERSION AC007274.3 GI:10801445
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 117309)
Sulston,J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
2 (bases 1 to 117309)
Mcpherson-Wagner,C., Maupin,R. and Krasucki,Peter.
The sequence of Homo sapiens BAC clone RP11-105L10
Unpublished
3 (bases 1 to 117309)
Waterston,R.H.
Direct Submission
Submitted (09-APR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 117309)
Waterston,R.
Direct Submission
Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 117309)
Waterston,R.
Direct Submission
Submitted (14-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 14, 2000 this sequence version replaced gi:5624109.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0105L10

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaltsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-305H21; the clone sequenced to the right is RP11-109F19, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-105L10; actual end is at base position 24605 of RP11-109F19.

FEATURES

Location/Qualifiers	Source
1..117309	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="Y"	
/map="Y"	
/clone="RP11-105L10"	
/clone_lib="RPCI-11"	
733..1035	
/rpt_family="Alu"	repeat_region
1787..1857	repeat_region
/rpt_family="MIR"	repeat_region
1967..2286	repeat_region
/rpt_family="Alu"	repeat_region
3460..3864	repeat_region
/rpt_family="MIR"	repeat_region
4654..4753	repeat_region
/rpt_family="MIR"	repeat_region
4917..5048	repeat_region
/rpt_family="MIR"	repeat_region
6100..6400	repeat_region
/rpt_family="Alu"	repeat_region
6920..7254	repeat_region
/rpt_family="L2"	repeat_region
7363..7578	repeat_region
/rpt_family="ERV"	repeat_region
7386..8065	repeat_region
/rpt_family="ERV1"	repeat_region
8066..8419	repeat_region
/rpt_family="ERV"	repeat_region
8762..8877	repeat_region
/rpt_family="ERV1"	repeat_region
8883..8945	repeat_region
/rpt_family="Alu"	repeat_region
8961..9178	repeat_region
/rpt_family="ERV1"	repeat_region
9185..10437	repeat_region





```
CDS
/db_xref="FLYBASE:FBgn0022943"
/join(4165..4212,4268..4485,4610..4808)
/gene="Cbp20"
/note="Cbp20 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012357"
/db_xref="FLYBASE:FBgn0022943"
/protein_id="AAF55496.1"
/db_xref="GI:7300337"
/translacion="MFASVELSYRDOHQFKSRSEORSLRDSCITLVVGNLSFYTTEE
QIHLEFSRCGDVIVIMGLDKYKTCGFCFVEYVYRSEAEAMRVNGFRLLDRILR
VWDAGFEVGRQYGRGKTGGQVDEYRTDYDAGGGYKLLSKKIAPNTDNR"
complement(join(4878..5331,5462..5596,5791..5935))
/gene="CG7217"
/product="CT22257"
/db_xref="FLYBASE:FBan0007217"
/db_xref="FLYBASE:FBgn0038570"
complement(<4878..>5935)
/gene="CG7217"
/map="90E3-90E4"
/db_xref="FLYBASE:FBan0007217"
/db_xref="FLYBASE:FBgn0038570"
complement(join(5002..5331,5462..5596,5791..5799))
/gene="CG7217"
/note="CG7217 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007217"
/db_xref="FLYBASE:FBgn0038570"
/protein_id="AAF55497.1"
/db_xref="GI:7300338"
/translacion="MVKGVSLFSDVLFEDSPANKINTGDLVNGKKVIIFGVPGATFP
GCSKTHLPGVSSADELKSQGVDEIVCVSNDFPVMASANGKEHGAAGKRVLLADPAG
GFTKALDVTIDLPPLGCVRSKRYSLVWENKVTLELVNVEPDCTGLSCSLANNIGKK"
complement(join(<6163..6507,6567..>6614))
/gene="CG7215"
/product="CT22253"
/db_xref="FLYBASE:FBan0007215"
/db_xref="FLYBASE:FBgn0038571"
complement(<6163..>6614)
/gene="CG7215"
/map="90E3-90E4"
/db_xref="FLYBASE:FBan0007215"
/db_xref="FLYBASE:FBgn0038571"
complement(join(6163..6507,6567..6614))
/gene="CG7215"
/note="CG7215 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007215"
/db_xref="FLYBASE:FBgn0038571"
/protein_id="AAF55498.1"
/db_xref="GI:7300339"
/translacion="MQTIKVLKGDCTIEVAPTSTILEYKHQIEAELOISATNOKLL
LLGRPLANEQTIASYPNIKEGTKLNLVWIKPLRDSILRGFRKHSEPLAERTNFM
ADPERKTNEQSLODLELADSVNRAT"
/join(7165..7303,9093..9534,10965..>11147)
/gene="CG7901"
/product="CT23882"
/db_xref="FLYBASE:FBan0007901"
/db_xref="FLYBASE:FBgn0038572"
/join(7165..7303,8920..9534,10965..>11147)
/gene="CG7901"
/product="CT2545"
/db_xref="FLYBASE:FBan0007901"
/db_xref="FLYBASE:FBgn0038572"
<7165..>11147
/gene="CG7901"
/map="90E3-90E5"
/db_xref="FLYBASE:FBan0007901"
/db_xref="FLYBASE:FBgn0038572"
/join(9142..9534,10965..11147)
/gene="CG7901"
/note="CG7901 gene product [alt 1]"
/codon_start=1
```

```
CDS
/db_xref="FLYBASE:FBan0007901"
/db_xref="FLYBASE:FBgn0038572"
/protein_id="AAF55499.1"
/db_xref="GI:7300340"
/translacion="MDNEALDPTIKSSTAATPAASETTTTAASSVVVETTTTAAAT
TASAEKNETTATNNSNTSGSISSSSNNIVIPASATNGIKESNSNLSTTTTAAAY
AAATVEGVAPAITSTIVTGGTTPPLSSLANLKUDNTPPYDAPPPPIISKVLNITGTP
IVRKERTSARYNASKNCELTALIPNESE"
/join(9142..9534,10965..11147)
/gene="CG7901"
/note="CG7901 gene product [alt 2]"
/codon_start=1
/db_xref="FLYBASE:FBan0007901"
/db_xref="FLYBASE:FBgn0038572"
/protein_id="AAF55500.1"
/db_xref="GI:7300341"
/translacion="MDNEALDPTIKSSTAATPAASETTTTAASSVVVETTTTAAAT
TASAEKNETTATNNSNTSGSISSSSNNIVIPASATNGIKESNSNLSTTTTAAAY
AAATVEGVAPAITSTIVTGGTTPPLSSLANLKUDNTPPYDAPPPPIISKVLNITGTP
IVRKERTSARYNASKNCELTALIPNESE"
/join(12280..12385,12562..13626,13976..14480,14769..14965,
15039..15373,15432..16135,16241..16677)
/gene="CG7913"
/product="CT23900"

mRNA
*
US-09-847-539a-6_COPY_59_86 (1-28) x AE003721 (1-224896)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13577 GAGGAGGAAGCTGACGAGGAGGACGCCGCGAGGATGCGCGCGGTCGCCGCC 13518

QY 25 AspAsnAlaAla 28
||| |||||

Db 13517 GATGCTGGGCG 13506

RESULT 16
SCOA14671
LOCUS SCOA14671 7686 bp DNA circular BCT 19-DEC-2001
DEFINITION Streptomyces coelicolor plasmid 2 (SCP2*).
ACCESSION AJ414671
VERSION AJ414671.1 GI:17974213
KEYWORDS ORF1; ORF2; ORF3; ORF4; ORF5; ORF6.
SOURCE Streptomyces coelicolor.
ORGANISM Streptomyces coelicolor
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (sites)
AUTHORS Wilkison,C.J., Hughes-Thomas,Z.A., Martin (nee Rowe),C.J.,
Bohm,I., Mironenko,T., Deacon,M., Wheatcroft,M., Wirtz,G.,
Staunton,J. and Leadlay,P.F.
Increasing the efficiency of heterologous promoters in
actinomycetes
J. Mol. Microbiol. Biotechnol. In press
REFERENCE 2 (bases 1 to 7686)
AUTHORS Wilkison,C.J.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-2001) Wilkison C.J., Department of Biochemistry,
University of Cambridge, 80 Tennis Court Rd., Cambridge, CB2 1GA,
UNITED KINGDOM
FEATURES
source location/Qualifiers
1..7686
/organism="Streptomyces coelicolor"
/plasmid="2 SCP2*"
/db_xref="taxon:1902"
1261..1713

CDS
```

\* 152667 153590: contig of 924 bp in length  
\* 153591 153670: gap of unknown length  
\* 153671 154319: contig of 649 bp in length  
\* 154320 154399: gap of unknown length  
\* 154400 154743: contig of 344 bp in length  
\* 154744 154823: gap of unknown length  
\* 154824 155422: contig of 599 bp in length  
\* 155423 155502: gap of unknown length  
\* 155503 155840: contig of 338 bp in length.

## FEATURES

source

1. .155840  
/organism="Drosophila melanogaster"  
/strain="y: cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="3"  
/map="90E1-90F10"  
/clone\_lib="BAC48113 (D522) RPT-98 48.1.13"  
/clone\_lib="RPT-98 (Roswell Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
pRAC3.6"

BASE COUNT 41735 a 34360 c 34802 g 41426 t 3517 others  
ORIGIN

## Alignment Scores:

Pred. No.: 3,23e+03 Length: 155840  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservatives: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AC006495 (1-155840)

QY 5 GluAlaLeuAlaSpGlnThrAspAlaLeuGlnSerGluCluAlaValValysAla 24  
||| |||||: ||||| :|||: ||||| |||  
Db 13996 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGAGGATCGCGCGCG 13937

QY 25 AspAsnAlaAla 28  
||| |||||

Db 13936 GATCGTCGCGCG 13925

## RESULT 15

AE003721/c

LOCUS Drosophila melanogaster genomic scaffold 14200001386035 section 46  
DEFINITION of 105, complete sequence.

ACCESSION AE003721 AE002708

VERSION AE003721.1 GI:7300335

KEYWORDS HTG.

SOURCE fruit fly.

## ORGANISM

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

AUTHORS

Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Anatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers J.H., Blazek R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor  
Miklos G.L., Abril J.F., Aghayani A., An H.J., Basu A.,  
Andrews Pfannkuch C., Baldwin D., Ballew R.M., Beeson K.Y.,  
Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y.,  
Benos P.V., Bertram B.P., Bhandari D., Bolshakov S., Borkova D.,  
Botchan M.R., Bouck J., Brokstein P., Brottier P., Burtis K.C.,  
Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de  
Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dletz S.M.,  
Doddson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C.,  
Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S.,  
Fleischmann W., Folsler C., Gabriellian A.E., Garb N.S.,  
Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z.,

Guan P., Harris M., Harris N.L., Harvey D., Helman T.J.,  
Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J.,  
Wel M.H., Ibegwam C., Jaiall M., Kalush F., Karpen G.H., Ke Z.,  
Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C.,  
Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A.,  
Li J., Li Z., Liang Y., Lin X., Liu X., Mattel B., McIntosh T.C.,  
McLeod M.P., McPherson D., Merkulov G., Miliushina N.V., Moberly C.,  
Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L.,  
Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K.,  
Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S.,  
Pollard J., Puri V., Reese M.G., Reinert K., Renington K.,  
Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kianos I.,  
Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C.,  
Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R.,  
Venter E., Wang A.H., Wang X., Wang Z.Y., Wasserman D.A.,  
Weinstock G.M., Weissbach J., Williams S.M., Woodage T.,  
Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yah R.F.,  
Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H.,  
Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M. and Venter J.C.  
The genome sequence of *Drosophila melanogaster*  
Science 287 (5461), 2185-2195 (2000)  
20196006  
2 (bases 1 to 224896)  
Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M. and Venter C.J.  
Direct Submission  
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
Location/Qualifiers  
1. .224896  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="3R"  
complement(join(1202..2465,2540..3150,3540..3806))  
/gene="CG7218"  
/product="CT22261"  
/db\_xref="FLYBASE:FBan0007218"  
/db\_xref="FLYBASE:FBgn0038569"  
complement(<1202..>3806)  
/gene="CG7218"  
/map="90E2-90E4"  
/db\_xref="FLYBASE:FBan0007218"  
/db\_xref="FLYBASE:FBgn0038569"  
complement(join(1245..2465,2540..3150,3540..3738))  
/gene="CG7218"  
/note="CG7218 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0007218"  
/db\_xref="FLYBASE:FBgn0038569"  
/protein\_id="AAF55495.2"  
/db\_xref="GI:10726596"  
/translation="MNATNSAGKKRQLRFRGDTGSRVEELHHQOQKQKAPLAQ  
DDVAATPTATPAAGSAQQLQSGTATCTNTFLPFRVFMVAALVSRVARIPLRS  
KIYSPMRIPLDERFWYQIMOCADSFYIHTFLPFRVFMVAALVSRVARIPLRS  
SGRLSPAECIDLLKGVITMTVTLMLLADVNRVHIIKQSIIKLYFYANLVEGD  
RLLSAGQDTIDALFWTATEPKNSKREHPCVLTHVLTLLYVFLSGLLNFONTCLNV  
AYNSAKGULTIMISNNFVELGSFKFKNNLQTLTCDVREPHLSVLTFLVVIQ  
TKREDFWSTQFCVMLPDCFAVLTEILIDWYKHAFITFENLPEISYREYTSLAYD  
MTQKQKHAFSDSLVARRMGFIPLPAVLIKATYATSFENLAAWLLFLAYLFA  
MGLRLCTICALGKACAKKEHQETERNSTPSSMTNPVPIGAAAPVSAATGQNHNN  
NNNNNNISIGSKPAQVTTLLTPPSAGHLDSKFNFSIASTSTPKKAVSDELDV  
TNSLELATALPSNSDVLDDVCLNEQVNTNTSSAEOVQOQDLVRSOPDLMLNN  
SGDVTSAKAKRATKRLPRTKRSSEPMFSWKEGGAAGTAGGNTQTQL."  
join(4128..4212,4268..4485,4610..4873)  
/gene="Cbp20"  
/product="CT23846"  
/db\_xref="FLYBASE:FBan0012357"  
/db\_xref="FLYBASE:FBgn0022943"  
<4128..>4873  
/gene="Cbp20"  
/note="CG12357"  
/map="90E3-90E4"  
/db\_xref="FLYBASE:FBan0012357"

mRNA

gene





/gene="GRL5"  
/note="Ca2+ binding protein; translated as a proprotein;  
Grl5p"  
/codon\_start=1  
/transl\_table=6  
/product="granule lattice protein 5 precursor"  
/protein\_id="AAC27989.1"  
/db\_xref="GI:3366655"  
/translation="MRVIAALLVIALVCOSMAVTSKSOAKLMEKINSKLEKSHLGR  
ALKNWTATIKGYQDLYDAFAALKNOLLNLDNENSLFETOTASHDSAVAQFNAD  
ISNYNGIINDAQDLNLSNTYQONLDQAQQLQDNDTALNAEAEALANEAALYQ  
VATETSYNADQVIGLAEKLOERASHYDNADLGSFVQIKNTFSAQKVTETATKGM  
NAKHQLEFVPVQAMQVKNNTSQSSIQTAIKALQDLQAYFQKTFSDLNVEYVFTQNI  
VOSTDGLNQITDILQNOVLPGEAQISSLQAOIQOVEDALALAQOQLDQAQNALDAE  
NAQWENVARHQALVDRIINSEYDLVTQADRIYRNAQAVNGSD"  
195..1154  
mat\_peptide

/gene="GRL5"

/note="Grl5p"

/product="granule lattice protein 5"

BASE COUNT 417 a 314 c 202 g 355 t

ORIGIN

## Alignment Scores:

Pred. No.:	20.5	Length:	1288
Score:	54.00	Matches:	12
Percent Similarity:	66.67%	Conservative:	4
Best Local Similarity:	50.00%	Mismatches:	8
Query Match:	41.86%	Indels:	0
DB:	3	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AF031321 (1-1288)

QY 2 ASPALALEUgLuAlaLeuAAlaAaspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21  
Db 420 GATGCTCAACAAAGCTCTTAAGACAACACTGATGCTCTCAATGCTGCCGAAGAGCTCTT 479

QY 22 VallysAlaAasp 25

Db 480 GCCAATGCTGAA 491

## RESULT 12

AC014470/c

LOCUS

DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.

AC014470

VERSION AC014470.1 GI:6436865

KEYWORDS HTG; HTGS\_PHASE2.

SOURCE fruit fly.

## ORGANISM

Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 32270)

AUTHORS Adams, M. and Venter, J.C.

TITLE Direct Submission

JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

COMMENT This sequence was identified as CDL:1021984 by the submitter.  
For further information on this sequence e-mail to fly@celera.com.

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

## FEATURES

source

Location/Qualifiers

1..32270

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT 9248 a 7110 c 6761 g 9151 t

ORIGIN

## Alignment Scores:

Pred. No.:	613	Length:	32270
------------	-----	---------	-------

Score:	54.00	Matches:	12
Percent Similarity:	66.67%	Conservative:	4
Best Local Similarity:	50.00%	Mismatches:	8
Query Match:	41.86%	Indels:	0
DB:	2	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC014470 (1-32270)

QY 5 GluAlaLeuAlaAaspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLysAla 24  
Db 26993 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGAGAGCGGTGCGGCC 26934

QY 25 AspAsnAlaAla 28  
Db 26933 GATGCTCGCGG 26922

## RESULT 13

AC009462/c

LOCUS

DEFINITION

Drosophila melanogaster, 151610 bp DNA linear INV 06-SEP-2001

BACR27G04, complete sequence.

AC009462

VERSION AC009462.6 GI:15451514

KEYWORDS HTG.

SOURCE fruit fly.

## ORGANISM

Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 151610)

## AUTHORS

Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,  
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,  
Rogers, Y., An, H., Baidwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,  
Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,  
Dodson, K., Dorsett, V., Dou, L.E., Doyle, C., Dreanek, D., Farfan, D.,  
Fierfiera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,  
Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J.,  
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., Moshrefi, A.,  
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,  
Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,  
Phuanavanong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,  
Shapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,  
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 3R, region 90E-90E

Unpublished

2 (bases 1 to 151610)

## JOURNAL

REFERENCE

## AUTHORS

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,  
Butenhoff, C., Champs, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomtan, M.A., Mazda, P.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclet, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,  
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.

Direct Submission

## JOURNAL

COMMENT

Submitted (23-AUG-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Sep 6, 2001 this sequence version replaced gi:13384332.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome  
shotgun and from subclones of this BAC and its neighboring clones.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

## FEATURES

source

1..151610

Location/Qualifiers

/organism="Drosophila melanogaster"

/strain="y; cn bw sp"

```

/db_xref="GI:16505984"
/translation="MKILVRISASTDYVYFLFMVKCDLNDDEIQAAIERNLVEYTG
MDADSVYDDGVCWHNGSCWVDDTMTVSDAHLERILGISTFE"
complement(658..663)
/feature="possible RBS"
gene
complement(719..1129)
/feature="HCM2.0004c"
CDS
complement(719..1129)
/feature="HCM2.0004c"
/feature="HCM2.0004c, hypothetical protein, len: 136 aa;
highly similar to TR:CA855229 (EMBL:AL117211), YPMT1.47C,
Yersinia pestis CO-92 hypothetical protein from plasmid
pMT1 (136 aa), fasta scores: E(): 0, 97.8% identity in 136
aa overlap and TR:O68762 (EMBL:AF074611), Y1116, Yersinia
pestis KIM5 hypothetical protein from plasmid pMT1 (136
aa), fasta scores: E(): 0, 97.8% identity in 136 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09871.1"
/db_xref="GI:16505985"
/translation="MKILYDAANALDVVDTELACGFPEPEWATOLREAIEMRAPEP
SEFADARQFIRMYAEVGPPTAQAMLLKYFKRAGENLPVDDTPHFAWRKFDV
IYTRMGSKRDVYVHLMHLDKRAVDTELEKFFPPA"
complement(1113..1484)
/feature="HCM2.0005c"
CDS
complement(1113..1484)
/feature="HCM2.0005c"
/feature="HCM2.0005c, hypothetical protein, len: 123 aa;
similar to TR:Q38665 (EMBL:X78401), ninX, Bacteriophage
P22 NinX protein (112 aa), fasta scores: E(): 8.5e-15,
45.6% identity in 114 aa overlap. Highly similar to
TR:CA855230 (EMBL:AL117211), YPMT1.48C, Yersinia pestis
CO-92 hypothetical protein from plasmid pMT1 (123 aa),
fasta scores: E(): 0, 87.8% identity in 123 aa overlap
and TR:Q92GF0 (EMBL:AF074611), Y1115, Yersinia pestis KIM5
hypothetical protein from plasmid pMT1 (123 aa), fasta
scores: E(): 0, 87.8% identity in 123 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09872.1"
/db_xref="GI:16505986"
/translation="MKNYSEMTDFEINCLVAEATGHRPLISQYWGKSGEGDYAVVA
IGPAGATFDWCNDPEDAWDIYRHRIGVIPARQPCWRAHRKVDSTPQHLIQPN
PFRAAMTVFLMQEKKHETV"
complement(1137..1141)
/feature="HCM2.0005c"
RBS
complement(1492..1497)
/feature="possible RBS"
RBS
complement(1492..1497)
/feature="possible RBS"
repeat_unit
1565..1621
/feature="repeat 1; 57 bp repeat"
gene
complement(1647..1649)
/feature="HCM2.0006c"
CDS
complement(1647..1649)
/feature="HCM2.0006c"
/feature="HCM2.0006c, possible lipoprotein, len: 280 aa;
shows weak similarity to TR:AAF10061 (EMBL:AE001907),
DR0482, Deinococcus radiodurans B-cell receptor associated
protein-related protein (328 aa), fasta scores: E():
4.9e-07, 23.4% identity in 265 aa overlap. Highly similar
to TR:CA855231 (EMBL:AL117211), YPMT1.49C, Yersinia pestis
CO-92 putative lipoprotein from plasmid pMT1 (276 aa),
fasta scores: E(): 0, 97.5% identity in 280 aa overlap and
TR:O68763 (EMBL:AF074611), Y1114, Yersinia pestis KIM5
hypothetical protein from plasmid pMT1 (276 aa), fasta
scores: E(): 0, 97.5% identity in 280 aa overlap. Contains
a possible N-terminal signal sequence and an appropriately
positioned PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
/codon_start=1

```

```

/transl_table=11
/product="putative lipoprotein"
/protein_id="CAD09873.1"
/db_xref="GI:16505987"
/translation="MKGLLAVALTACTMGLTGCDRVRVPGYVGIKYNKLGEDKGIGE
VVGVGROWTGLNTELYTPTPKQMKTYDEPFTFQMSDGTAGHKIGVAYLVNRDKVTT
PVQYIRKGVDDITDIDLRKSLNRIASRMTDTTFIDGKASLNDLNAKDKIOAEMS
FQYIEVLSLVGKPDYPTVIESINAKVTANQTKLQRRQVEQRKAEANMLRQQAEG
EADGEADAIRKRAQAEADAIKLRGEALRQNPVMELEAIKNGOLPOYNTGANTPE
IALK"
complement(2427..2459)
misc_feature
/feature="HCM2.0006c"
/feature="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
gene
complement(2493..2693)
CDS
/feature="HCM2.0007c"
/feature="HCM2.0007c"
/feature="HCM2.0007c, possible membrane protein, len: 66 aa;
highly similar to TR:CA852372 (EMBL:AL117211), YPMT1.49AC,
Yersinia pestis CO-92 hypothetical protein from plasmid
pMT1 (66 aa), fasta scores: E(): 3.5e-25, 97.0% identity
in 66 aa overlap. Contains hydrophobic, possible
Alignment Scores:
Pred. No.: 718 Length: 106516
Score: 57.00 Matches: 13
Percent Similarity: 68.18% Conservative: 2
Best Local Similarity: 59.09% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6_COPY_59_86 (1-28) x STYPPHCM2 (1-106516)
QY 6 AlaLeuAlaAspGluThrAspAlaLeuGlnSerGluGluAlaAlaVallylsaaap 25
Db 39586 GCGGAGCATCAACAACAGCTGCATTACAAAGTGACAGGTTCACGCCCATCGACAC 39527
QY 26 AsnAla 27
Db 39526 TCCGCG 39521
RESULT 11
AF031321
LOCUS
DEFINITION AF031321 1288 bp mRNA linear INV 31-JUL-1998
Tetrahymena thermophila granule lattice protein 5 precursor (GRL5)
ACCESSION AF031321
VERSION AF031321.1 GI:3366654
KEYWORDS
SOURCE Tetrahymena thermophila.
ORGANISM Tetrahymena thermophila
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE 1 (bases 1 to 1288)
AUTHORS Verbsky,J.W. and Turkewitz,A.P.
TITLE proteolytic processing and Ca2+ binding activity of dense-core
vesicle polypeptides in Tetrahymena
JOURNAL Mol. Biol. Cell (1997) In press
REFERENCE 2 (bases 1 to 1288)
AUTHORS Verbsky,J.W. and Turkewitz,A.P.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-1997) Molecular Genetics and Cell Biology,
University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA
FEATURES
source
1..1288
/organism="Tetrahymena thermophila"
/strain="B1686"
/db_xref="taxon:5911"
gene
1..1288
/feature="GRL5"
/allele="wt"
24..1157
CDS

```



Diane E. Taylor are with the University of Alberta Department of Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G 2H7 and the Canadian Bacterial Diseases Network. All other authors are with Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capability are available at ARCB's World Wide Web site. (URL: <http://www.astra-boston.com/hpylori>).

## FEATURES

source	1. 12526 /organism="Helicobacter pylori J99" /strain="J99" /db_xref="taxon:85963" 114. 1244 /gene="jhp1241" 114. 1244 /gene="jhp1241" /note="similar to H. pylori 26695 gene HP1321" /codon_start=1 /transl_table=11 /product="putative" /protein_id="AAD06826.1" /db_xref="GI:4155853" /translation="MRYDCLNPITPIRKFMPISLCHPFGFETPKPEPALNPLIKAHSDKICLLGPKHSGKTFITFALKAKGKNPYNINCMDRLNQNLISWLLRWLEKKMDLLIDIRERLDFSLPKPLKIVLILNYLSHTAPNFSLCYALGNFKEYISFPKNTPKNTLNRFLRDNALDSLPTENOEKILKKQENIOLAFQAYAPLMAKICYSQSKFVSAPYLYTOLKELSTKDTYLLKLALEKORILIELAPNFENHKTLYLCPALPYSYTPSPSLNVFNWVLELYKQPNYELVAHNGIFILRNPTSKALIAHAFPTPFLEKQLLWCHKRGFKIAVVSINAPISANNPPYKHLNFIIDFSLOISILV" 1509. .2087 /gene="jhp1242" 1509. .2087 /gene="jhp1242" /note="similar to H. pylori 26695 gene HP1322" /codon_start=1 /transl_table=11 /product="putative" /protein_id="AAD06827.1" /db_xref="GI:4155854" /translation="MNALKLSFALLSLGLFAQTVHAOHLKDTINYPDWLKNLEDPKKNPNQIVGSRASISGRNDEFSYNIPIYDDRLPPEKNAEEIALRARNAYSTLESALLDLIAFYKRIKIDLSASATNLSALNLKALFHL" complement(2061. .2690) /gene="rnhb" /note="jhp1243" complement(2061. .2690) /gene="rnhb" /note="similar to H. pylori 26695 gene HP1323" /codon_start=1 /transl_table=11 /product="RIBONUCLEASE HII" /protein_id="AAD06828.1" /db_xref="GI:4155855" /translation="MGCVMTLGIIDEAGRCGLAGSLFVAGVVCNEKIALEFLKMLKDSKILSPKRRFLEDKIRTHGVGFVYVKKSAANEIDHGLGACLKLAIEIIVENGCSIA NEIKDGNATFGLNKRYPNIQTIIKDETIAQIAMASVLAKASKDREMLHALFKEYGWKNCGGYTKOHEATNKLGAATPFHRHSFTLKNRLNPKLLEVEQRLV" complement(2706. .2900) /gene="jhp1244" complement(2706. .2900) /gene="jhp1244" /note="similar to H. pylori 26695 gene HP1324" /codon_start=1 /transl_table=11 /product="putative" /protein_id="AAD06829.1" /db_xref="GI:4155856" /translation="MHDNDLNLVLESGSKHAKLTGRVHHLTIPSSPSDRKSVKNEFELTEFIKKLCGNRLTHGAF" complement(2981. .4372) /gene="fumc"
gene	/note="jhp1245" complement(2981. .4372) /gene="fumc" /note="similar to H. pylori 26695 gene HP1325" /codon_start=1 /transl_table=11 /product="FUMARASE" /protein_id="AAD06830.1" /db_xref="GI:4155857" /translation="MOPRIEHTMGEIKVDSDOYWGAAOTORSLENFKIGTEKPKPELIFCAPAKLRSLAVYVNHKLCKLSLEKSOAIKACDILKGEICGCEFFLAINGTSGTQTNMNLNEVIANKATEILGNFREKKLIHPNDVNNWSQNSNDTFTTAMHVVSVEIETHKLLPSLENLIKTFKESQOFKEIVKIGTRTHLQDAPTLTLGQEFSGYASMLEHSKQOILLES EHLRELAIGGTAVGTGLNHLKSELKVAEELSQFSGVKFTSAPNKFTHALTSHDAIAYAHGAFKALAAANLMIANDIRWLASGPRCGELNIPENPGSSIMPQKVNPTQCEAMTVAVQVAGNDTAIGIAASQGNFELNPKPVIINYFLOSIRLLSDSMESFNHCASGIEP NREKDIYLLHHSLSMLVTALNPHVGYENAAKIAHNAHKKGISLAKESALELKLSSAEFDFQFVVPKMKIGPKA" 4547. .4924 /gene="jhp1246" 4547. .4924 /gene="jhp1246" /note="similar to H. pylori 26695 gene HP1326" /codon_start=1 /transl_table=11 /product="putative" /protein_id="AAD06819.1" /db_xref="GI:4155846" /translation="MRKLAALFLVSALGVNSLNAWEQTLKANDELYKIKSVGNPIKGDNTVLSPTLKAGKALEKAIKVRQFMPMPGMPAMKEMAQVSEKNGLYEAKTNLSMNGTWOVRVDIKSKEGQVYRTKSLDL" 4929. .6167 /gene="jhp1247" 4929. .6167 /gene="jhp1247" /note="similar to H. pylori 26695 gene HP1327" /codon_start=1 /transl_table=11 /product="putative" /protein_id="AAD06820.1" /db_xref="GI:4155847" /translation="MLSFISAFDKRGVSIRLLTALLLLLSGLAKDLEIOTFFVAKYLSKNQIQALQPIDALDSQEKVSKWDPILYLGYNNAVNSDFRDLSTLQNMNSLGLSKQYDLNKKLTQSKMINLEKQKILEKTKQQLINMINGIENYKNOQEIELNTA IKNLENTLYQANHSSSPDLIAITAKLEILKSLLEKNDLEVALSSSHYSMGELTFKFN EILISAPKNEFNNEQELHNISATNYDIAIARLDEEKAQKIDITLAKKSFLEDINVTGVYFRSKOYVWYDMESVALSIPLPLYGKQAKLVEQKKESLAFKSEVENAKNKRHLAL KLLKLETLOKNLESINKIIEKQEKIAQIYALDLKTNQDYNAYNALNDKITIQTITOL ETLSALNSAYLSLQNLKGL" 6164. .7189 /gene="czcb_2" /note="jhp1248" 6164. 7189 /gene="czcb_2" /note="similar to H. pylori 26695 gene HP1328" /codon_start=1 /transl_table=11 /product="CATION EFFLUX SYSTEM PROTEIN" /protein_id="AAD06821.1" /db_xref="GI:4155848" /translation="MKRLLLALALFFLSLCTNAQEIKETQETKKTKETKSQTRFNIS TTKYIEKFSQSRRYALALPEQNEALIFSQTRFDGYVEKLYANKYTPPIKGDRLLSV YSPELAGVQSSELLSKFNQOGAIKELKMLIGLNFSTIEKTIISSHRVQNEITYSRF NGVIFKSPDLNCGSFTKGOELFKIIDLSRLWALVKVNOEDLEFLANTHQAILFVFG VKGQAITLENINPILNAQDKMLEARENVPNLKLLYVNPMAQVVEIFHKPKKILEK EAVLIRGKGAIVFKKDDFGLSPLEIKAVRUSDGSTEILEGLKAGEEVANALFVLIDAD AQNNGDY" 7190. .10297 /gene="czca_2" /note="jhp1249" 7190. 10297 /gene="czca_2" /note="similar to H. pylori 26695 gene HP1329"
gene	
CDS	

```

/translation="MNPEKATHCKEIRKILKALHNDKDLNLYLESGSKHAKLTSCAN
HLIIPSSDRKSVNFEKELTEFKRLRENTHITETCE"
complement(2949..4340)
/gene="Hp1325"
CDS
complement(2949..4340)
/gene="Hp1325"
/notes="similar to SP:P05042 GB:X00522 GB:X04065 PID:41513
GB:U00096 percent identity: 63.68; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="fumarase (fumC)"
/protein_id="AAD08368.1"
/db_xref="GI:2314492"
/translation="MQFRIEDHTWGEIKVNDQYWGCAQTORSLNFKIGTEKMPKELI
GAFALKSLAVNHNKLOKLSKQAIKACDCILKGLGCEFFLAIWOTGSGYOTN
MNLNIVANKATEIILGNPFREKKLIHPNDVNMSSNDTFTAMHIVSVLTHIRLL
PSLENLTKTFKESQOFKEIVKIGTHIQDAPLTLGQEFSGYASMLHSHKQOILLES
EHURELATGAVGTGLNAHKELSEKVAELSOFSGVFVSNPKFHALTSHDAIYA
HGAFAALANLWKIANDIRWLASGPRCGLGELINIPENEPGSSIMPCKVNPITCEANTM
VAVQVMDNTAIGIASQGNFELNFKPVIYINFLQSLRLSDSHESNHNHCAAGIEP
NREKIDYILHSHLMTVALNPHVGTENAAKIAKNAHKKGISLSEKALELKLUSAEDFD
KPVVPEKMGPKRA"
4515..4892
/gene="Hp1326"
CDS
4515..4892
/gene="Hp1326"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region Hp1326"
/protein_id="AAD08376.1"
/db_xref="GI:2314500"
/translation="MKKLAALFLVSLVGMVGLNABEQTLKANDLEVKIKSVGNPKGD
NFIILSPILKGALEKAIYRVQFMPEMPGMPANKEMAQVSEKNGLYEAKTNLSMNGT
WQVRVDIKSEKQVYRAKTSLDL"
4897..6135
/gene="Hp1327"
CDS
4897..6135
/gene="Hp1327"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region Hp1327"
/protein_id="AAD08377.1"
/db_xref="GI:2314501"
/translation="MLSFISAFDKRGVSIRLTALLLFLSLGAKDLEIQSFVAKVLS
KNQIQALQEQIDALSSQEKVKVDNPIILYLYNNANVSDFRLDSTLMQNSLGLS
QKVDLNGKRLQDSQMDLEKQKILELKKTKOOLALSLMINGIENKKNQOETELKTA
INKNLNTYQANHSSPNLTAIAKLEILKSQLEIKNNLEALSASHYSMGELAFKN
ELLSAPKNFENRREOELHNTSATNYDIAIARLDEKSKODITLAKKSFLEDVNTGV
YFRSKQYNDYMFIALSIPLIYKQAKLVEQKKESLVEKSEVENTKNTKTHLAL
KLLKLETLQKNLESINKIKONEKIAQIAYLDLKSNGDYNAYNAFNDKITQITQL
ETLSALNSTYLSLONKGLG"
6132..7148
/gene="Hp1328"
CDS
6132..7148
/gene="Hp1328"
/notes="similar to GB:M26073 SP:P13510 PID:141927 percent
identity: 28.85; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="cation efflux system protein (czca)"
/protein_id="AAD08369.1"
/db_xref="GI:2314493"
/translation="MKRIILWALILFFSLPFLANAQTOEIKKTEAKSOTRPNISTTK
VLEKFSOSRRYALLENPEALIFSQTRFDGVEKLYANKYTTIKKGRDLLSYSP
VIEQAQSLKSLKFNQOQVGAIKERKLIGLNSSEIKTISSHKQVNETIYSHFNGI
IFKSPDLNKGFSFKQGBLFOIIDLQWALWLVKNQOEDLEFLKNTHAILPVEKIG
EQEITLENINPINKEDKMLEARENVPNVKQIYYPNMFAQVEIFQKPKMKILPREAV
```

```

LKGGKATVFKKDDFGLSPLEIKAVRLSDGSEYIEGLKAGEEVANNALFLVLDADQN
NGDY"
7149..10256
/gene="Hp1329"
CDS
7149..10256
/gene="Hp1329"
/notes="similar to GB:M26073 SP:P13511 PID:141928 percent
identity: 31.32; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="cation efflux system protein (czca)"
/protein_id="AAD08370.1"
/db_xref="GI:2314494"
/translation="MIEKIIDLVSNNKLLTTLVLIFLASLWAKSVRLDALDPLSP
AQVVOITYPNQSPKIVQEVITYPLVSTFMSIANITDVRGSISSGLLIYIFKDGVN
LYWADRDLVEQLNRVSNLPKDAKVEIGSDSTIGWAYQYALSQSKNSDLVLQDFY
YVALGLGVSEVASGVGFVADYEVTQLQNDLIRYNLSLEQVANAIKNSNDTGGV
ILENGFEKIIIRSHGVIQSLKDLLEIIVKKEGAIPLKINDIASVRLTPKPRCAANLNG

Alignment Scores:
Pred. No.: 74.8 Length: 12496
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 1 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AE000634 (1-12496)

Qy 4 LeuGLuAlaAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23
Db 5041 ATACAAGCCTTACAGAGCAAAATGACGCTTTAAGTCTCTCAAGAAAAGTCGTAGCAAG 5100

Qy 24 AlaAspAsn 26
Db 5101 TGGGATAAC 5109

RESULT 9
AE001548
LOCUS
DEFINITION
Helicobacter pylori, strain J99 section 109 of 132 of the complete
genome.
ACCESSION
AE001548 AE001439
VERSION
AE001548.1 GI:4155845
KEYWORDS
Helicobacter pylori J99.
SOURCE
Helicobacter pylori J99
ORGANISM
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
Helicobacter.
REFERENCE
1 (bases 1 to 12526)
Alm,R.A., Ling,L.S., Moir,D.T., King,B.L., Brown,E.D., Dolg,P.C.,
Smith,D.R., Noonan,B., Guild,B.C., deJonge,B.L., Carmel,G.,
Tummino,P.J., Caruso,A., Uria-Nickelsen,M., Mills,D.M., Ives,C.,
Gibson,R., Merberg,D., Mills,S.D., Jiang,Q., Taylor,D.E.,
Vovis,G.F. and Trust,T.J.
Genomic-sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori
Nature 397 (6715), 176-180 (1999)
99120557
Erratum: [published erratum appears in Nature 1999 Feb
25;397(6721):719]]
2 (bases 1 to 12526)
King,B.L., Alm,R.A. and Trust,T.J.
Direct Submission
Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney
Street, Cambridge, MA 02139, USA
Address all correspondence to: hp@arch.us.astro.com or Richard
A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,
MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith,
Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,
Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics
Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and
```

[illegible]









ORGANISM Streptococcus pyogenes  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 468)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the  
bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL  
MEDLINE 99269061  
PUBMED 10336419  
REFERENCE 2 (bases 1 to 468)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular  
Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden  
FEATURES  
source  
1..468  
/organism="Streptococcus pyogenes"  
/strain="KMI3"  
/db\_xref="taxon:1314"  
mat\_peptide  
1..>468  
/product="GRAB"  
misc\_feature  
1..174  
/note="Region: alpha2-macroglobulin-binding A domain"  
CDS  
1..>468  
/note="grab"  
/codon\_start=1  
/transl\_table=11  
/product="GRAB precursor"  
/protein\_id="AAD26340.1"  
/db\_xref="GI:4589083"  
/translation="VDSPIEQPRIIPNGGTLNLGNAPKALRNEPRAIDELKKQA  
IEDKEATTAEASDSDALEALADOTDALQSEAAVVKADNAASDALEALADOTDALQS  
EEAEVQSDNAASDAWEKAATPIALDVKTKDTPVKRKEQRQNTLPTTGE"  
175..342  
/note="Region: repeat motif"  
misc\_feature  
343..468  
/note="Region: cell-wall attachment"  
BASE COUNT 178 a 91 c 100 g 99 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7 61e-12 Length: 468  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AFI24401 (1-468)  
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
Db 175 TCAGATGCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAAATCAGAAGAGCTGCG 234  
QY 21 ValVallyAlaAspAsnAlaAla 28  
Db 235 GTTGTAAAGCGGATAACGCTGCT 258  
RESULT 2  
AFI24400  
LOCUS AFI24400 469 bp DNA linear BCT 14-AUG-2000  
DEFINITION Streptococcus pyogenes strain AFI GRAB precursor, gene, partial  
cfs.  
ACCESSION AFI24400  
VERSION AFI24400.1 GI:4589080  
KEYWORDS Streptococcus pyogenes.  
SOURCE Streptococcus pyogenes  
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
REFERENCE 1 (bases 1 to 469)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the  
bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL  
MEDLINE 99269061  
PUBMED 10336419

TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the  
bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL  
MEDLINE 99269061  
PUBMED 10336419  
REFERENCE 2 (bases 1 to 469)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular  
Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden  
FEATURES  
source  
1..469  
/organism="Streptococcus pyogenes"  
/strain="AP1"  
/db\_xref="taxon:1314"  
mat\_peptide  
1..>469  
/product="GRAB"  
misc\_feature  
1..174  
/note="Region: alpha2-macroglobulin-binding A domain"  
CDS  
1..>469  
/note="grab"  
/codon\_start=1  
/transl\_table=11  
/product="GRAB precursor"  
/protein\_id="AAD26339.1"  
/db\_xref="GI:4589081"  
/translation="VDSPIEQPRIIPNGGTLNLGNAPKALRNEPRAIDELKKQA  
IEDKEATTAEASDSDALEALADOTDALQSEAAVVKADNAASDALEALADOTDALQS  
EEAEVQSDNAASDAWEKAATPIALDVKTKDTPVKRKEQRQNTLPTTGE"  
175..342  
/note="Region: repeat motif"  
misc\_feature  
343..469  
/note="Region: cell-wall attachment"  
BASE COUNT 178 a 90 c 100 g 101 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7 63e-12 Length: 469  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AFI24400 (1-469)  
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
Db 175 TCAGATGCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAAATCAGAAGAGCTGCG 234  
QY 21 ValVallyAlaAspAsnAlaAla 28  
Db 235 GTTGTAAAGCGGATAACGCTGCT 258  
RESULT 3  
AFI24402  
LOCUS AFI24402 717 bp DNA linear BCT 14-AUG-2000  
DEFINITION Streptococcus pyogenes strain KTL9 GRAB precursor, gene, partial  
cfs.  
ACCESSION AFI24402  
VERSION AFI24402.1 GI:4589084  
KEYWORDS Streptococcus pyogenes.  
SOURCE Streptococcus pyogenes  
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
REFERENCE 1 (bases 1 to 717)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the  
bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL  
MEDLINE 99269061  
PUBMED 10336419

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:49:23 ; Search time 274.909 Seconds  
(without alignments)  
2131.407 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALEALADQTDALQSEEAHVVKADNAA 28

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framep2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_093105\_4992/app\_query.fasta\_1.526  
-DB=GenEmbl -QFMT=fastap -SUFFIX=std.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=dot -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=score -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539.cgn\_1\_1182.etrnat\_10102002\_093105\_4992 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pi.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129	100.0	468	1	AF124401 Streptoco
2	129	100.0	469	1	AF124400 Streptoco
3	129	100.0	717	1	AF124402 Streptoco
4	129	100.0	804	1	AF124403 Streptoco
5	129	100.0	832	1	AF124399 Streptoco
6	129	100.0	10029	1	AE006573 Streptoco
7	57.5	44.6	3838	3	PFASA7 M10130 Plasmodium
8	57	44.2	12496	1	AE000634 Helicobac
9	57	44.2	12526	1	AE001548 Helicobac
10	57	44.2	106516	1	AE001548 Helicobac
11	54	41.9	1288	3	AF031321 Tetrahyme
12	54	41.9	32270	2	AC014470 Drosophil
13	54	41.9	151610	2	AC009462 Drosophil
14	54	41.9	155840	2	AC006495 Drosophil
15	54	41.9	224896	3	AE003721 Drosophil
16	53	41.1	7686	1	SC0414671 Streptomy
17	53	41.1	31317	1	SCSCP21 Streptomy
18	53	41.1	117309	9	AC007274 Homo sapi
19	53	41.1	173772	2	AC087862 Homo sapi
20	52.5	40.7	12411	1	AE004659 Pseudomon
21	52	40.3	12373	3	DME17920 Drosophila
22	52	40.3	16135	3	AF096897 Drosophil
23	52	40.3	17593	3	AF096896 Drosophil
24	52	40.3	36993	2	AC018007 Drosophil
25	52	40.3	133155	2	AC011910 Drosophil
26	52	40.3	169427	3	AC007416 Drosophil
27	52	40.3	281993	3	AE003619 Drosophil
28	52	40.3	303100	1	RME503646 Rhizobium
29	51.5	39.9	850	33	AC056141 Giardia
30	51.5	39.9	2888	1	ATU1522A Agrobacteri
31	51.5	39.9	12976	3	AC001654 Drosophil
32	51.5	39.9	32608	1	SCBAC36F5 Streptomy
33	51.5	39.9	35031	2	AC012724 Drosophil
34	51.5	39.9	110000	3	Continuation (3 of
35	51.5	39.9	170801	3	AC095014 Drosophil
36	51.5	39.9	190642	3	AC095015 Drosophil
37	51.5	39.9	309357	3	AE003673 Drosophil
38	51	39.5	2151	3	AY061146 Drosophil
39	51	39.5	12423	1	AF005155 Halobacte
40	51	39.5	14941	3	AF072878 Tetrahyme
41	51	39.5	20217	1	AE000047 Mycoplasma
42	51	39.5	30561	1	SC1F2 Streptomy
43	51	39.5	96288	3	AC005749 Drosophil
44	51	39.5	105823	2	AC019871 Drosophil
45	51	39.5	135509	8	AC037426 Oryza sat

ALIGNMENTS

RESULT 1

AF124401

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

AF124401 Streptococcus pyogenes strain K1L3 GRAB precursor, gene, partial cds.  
AF124401.1 GI:4589082  
Streptococcus pyogenes.

XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
 CC the probes of the invention.

XX SQ Sequence 544 BP; 124 A; 113 C; 135 G; 172 T; 0 other;

Alignment Scores:  
 Pred. No.: 85.9 Length: 544  
 Score: 45.50 Matches: 13  
 Percent Similarity: 55.56% Conservative: 2  
 Best Local Similarity: 48.15% Mismatches: 7  
 Query Match: 35.27% Indels: 5  
 DB: 22 Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AAK36405 (1-544)

Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16

Db 290 GACATTTGGACAGACTGCAGCTCAGGTGGATGCCCTTGACGGGCCATCTTCAACAAATG 349

Qy 17 GluGluAlaAlaValVallys 23

Db 350 GAGCAGGAAGCCCTGTGTGAAG 370

Search completed: October 13, 2002, 05:18:52  
 Job time : 1037.53 secs

```
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX PS Claim 1; SEQ ID NO 10489; 639pp + sequence listing; English.
XX
XX CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 544 BP; 124 A; 113 C; 135 G; 172 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 85.9 Length: 544
XX Score: 45.50 Matches: 13
XX Percent Similarity: 55.56% Conservative: 2
XX Best Local Similarity: 48.15% Mismatches: 7
XX Query Match: 35.27% Indels: 5
XX DB: 22 Gaps: 1
XX
XX US-09-847-539A-6_COPY_59_86 (1-28) x ABA62184 (1-544)
XX
XX QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16
XX Db 290 GACATTTTGGAAACAGACTGCAGTGGATGCTTGCAGGCCATCTTCACAAATG 349
XX
XX QY 17 GluGluAlaAlaValValLys 23
XX Db 350 GAGCAGGAAGCCCTGGTGAAG 370
XX
XX RESULT 39
XX AAK10508
XX ID AAK10508 standard; DNA; 544 BP.
XX
XX AC AAK10508;
XX
XX DT 05-NOV-2001 (first entry)
XX
XX DE Human brain expressed single exon probe SEQ ID NO: 10499.
XX
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer; ss.
XX
XX OS Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX DR
```

```
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX PS Example 4; SEQ ID NO: 10499; 650pp + Sequence Listing; English.
XX
XX CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX SQ Sequence 544 BP; 124 A; 113 C; 135 G; 172 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 85.9 Length: 544
XX Score: 45.50 Matches: 13
XX Percent Similarity: 55.56% Conservative: 2
XX Best Local Similarity: 48.15% Mismatches: 7
XX Query Match: 35.27% Indels: 5
XX DB: 22 Gaps: 1
XX
XX US-09-847-539A-6_COPY_59_86 (1-28) x AAK10508 (1-544)
XX
XX QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16
XX Db 290 GACATTTTGGAAACAGACTGCAGTGGATGCTTGCAGGCCATCTTCACAAATG 349
XX
XX QY 17 GluGluAlaAlaValValLys 23
XX Db 350 GAGCAGGAAGCCCTGGTGAAG 370
XX
XX RESULT 40
XX AAK36405
XX ID AAK36405 standard; DNA; 544 BP.
XX
XX AC AAK36405;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 10962.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX OS Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 10962; 658pp + Sequence Listing; English.
XX
XX PS
```

```
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX XX
XX PS Example 4; SEQ ID NO: 23882; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX SQ Sequence 234 BP; 61 A; 55 C; 61 G; 57 T; 0 other;
Alignment Scores:
Pred. No.: 29.6 Length: 234
Score: 45.50 Matches: 13
Percent Similarity: 55.56% Conservative: 2
Best Local Similarity: 48.15% Mismatches: 7
Query Match: 35.27% Indels: 5
DB: 22 Gaps: 1
US-09-847-539A-6_COPY_59_86 (1-28) x AAK49325 (1-234)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16
Db 154 GACATTTTGGACAGACTGCAGCTCAGGTGGATGCTTGCAGGGCCATCTTCAACAAATG 213
Qy 17 GluGluAlaAlaValVallys 23
Db 214 GAGCAGGAAGCCCTGGTGAAG 234
RESULT 37
AAI55172
ID AAI55172 standard; DNA; 234 BP.
XX AC AAI55172;
XX DT 17-OCT-2001 (first entry)
XX DE
XX DE Probe #23858 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX XX
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
```

```
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-48897/53.
XX XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX XX
XX PS Claim 25; SEQ ID NO 23858; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 234 BP; 61 A; 55 C; 61 G; 57 T; 0 other;
Alignment Scores:
Pred. No.: 29.6 Length: 234
Score: 45.50 Matches: 13
Percent Similarity: 55.56% Conservative: 2
Best Local Similarity: 48.15% Mismatches: 7
Query Match: 35.27% Indels: 5
DB: 22 Gaps: 1
US-09-847-539A-6_COPY_59_86 (1-28) x AAI55172 (1-234)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16
Db 154 GACATTTTGGACAGACTGCAGCTCAGGTGGATGCTTGCAGGGCCATCTTCAACAAATG 213
Qy 17 GluGluAlaAlaValVallys 23
Db 214 GAGCAGGAAGCCCTGGTGAAG 234
RESULT 38
ABA62184
ID ABA62184 standard; DNA; 544 BP.
XX AC ABA62184;
XX DT 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #10489.
XX DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX XX
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
```

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

PN WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -

XX Claim 4; SEQ ID NO 22983; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 234 BP; 61 A; 55 C; 61 G; 57 T; 0 other;

Alignment Scores:

Pred. No.:	29.6	Length:	234
Score:	45.50	Matches:	13
Percent Similarity:	55.56%	Conservative:	2
Best Local Similarity:	48.15%	Mismatches:	7
Query Match:	35.27%	Indels:	5
DB:	22	Gaps:	1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x ABA74678 (1-234)

OY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16

DB 154 GACATTTTGGACAGACTGCAGCTCAGGTGGATGCCTTGCAGGGCCATCTTCACAAATG 213

OY 17 GluGluAlaAlaValVallys 23

DB 214 GAGCAGGAAGCCCTGGTGAAG 234

RESULT 35

AAK23154

ID AAK23154 standard; DNA; 234 BP.

XX AAK23154;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID NO: 23145.

XX Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer; ss.

OS Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

PT brains -

XX Example 4; SEQ ID NO: 23145; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.

XX Sequence 234 BP; 61 A; 55 C; 61 G; 57 T; 0 other;

Alignment Scores:

Pred. No.:	29.6	Length:	234
Score:	45.50	Matches:	13
Percent Similarity:	55.56%	Conservative:	2
Best Local Similarity:	48.15%	Mismatches:	7
Query Match:	35.27%	Indels:	5
DB:	22	Gaps:	1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAK23154 (1-234)

OY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16

DB 154 GACATTTTGGACAGACTGCAGCTCAGGTGGATGCCTTGCAGGGCCATCTTCACAAATG 213

OY 17 GluGluAlaAlaValVallys 23

DB 214 GAGCAGGAAGCCCTGGTGAAG 234

RESULT 36

AAK49325

ID AAK49325 standard; DNA; 234 BP.

XX AAK49325;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 23882.

XX Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.



```
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX PT mutation point of a gene, measuring expression of a gene, analysing
XX PT expression profile or pattern of a gene and identifying homologous gene
XX
XX
XX Disclosure: SEQ ID NO: 7060; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX CC are useful for identifying the mutation point of a gene derived from a
XX CC mutant of coryneform bacterium, measuring expression amount and
XX CC analysing the expression profile or expression pattern of a gene derived
XX CC from coryneform bacterium, and identifying a homologue of a gene derived
XX CC from coryneform bacterium. Coryneform bacteria are useful for producing
XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX CC particularly L-lysine. The present sequence is a nucleic acid described
XX CC in the exemplification of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC European Patent Office.
XX
XX SQ Sequence 349980 BP; 80289 A; 91081 C; 97378 G; 81232 T; 0 other;

Alignment Scores:
Pred. No.: 2.41e+05 Length: 349980
Score: 46.00 Matches: 9
Percent Similarity: 62.50% Conservative: 6
Best Local Similarity: 37.50% Mismatches: 9
Query Match: 35.66% Indels: 0
DB: 22 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AAH68525 (1-349980)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLys 23
Db 225435 ATTGAGCCTTACGCACCTGAGGACGAGCGTGAAGAGGGCGAGTCCGCAATCTACAAG 225494
QY 24 AlaAspAsnAla 27
Db 225495 TTTCGAGAGCGCT 225506

RESULT 33
AAAB1490
ID AAAB1490 standard; DNA; 1437668 BP.
XX
XX AC AAAB1490;
XX
XX 04-DEC-2000 (first entry)
XX
XX N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
XX KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX KW Meningococcus B; MenB; ds.
XX
XX OS Neisseria meningitidis.
XX
XX PN WO200022430-A2.
XX
XX 20-APR-2000.
XX
XX 08-OCT-1999; 99WO-US23573.
XX

09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
XX PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
XX PI Rappuoli R, Pizza M;
XX
XX WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
XX PT used in the diagnosis and treatment of N. meningitidis infection and
XX PT other Neisserial infections, for example, N.gonorrhoea -
XX
XX Claim 7; Page 866-1272; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
XX CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
XX CC represent specifically claimed Neisseria meningitidis genomic DNA
XX CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
XX CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
XX CC AA81259 and AAA81304 to AAA81321 represent PCR primers used in the
XX CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
XX CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
XX CC sequences, which are all used in the exemplification of the present
XX CC invention. The nucleic acid sequences, protein sequences, and antibodies
XX CC against them, can be used in the manufacture of a composition. The
XX CC composition can be used as a medicament (or in the manufacture of a
XX CC medicament) for treating, preventing or diagnosing infection due to
XX CC Neisserial bacteria. For example, some of the identified proteins could
XX CC be components of vaccines against Meningococcus B; against all serotypes;
XX CC and/or against all pathogenic Neisseriae. Identification of sequences
XX CC from the bacterium will also facilitate production of biological probes,
XX CC particularly organism-specific probes. Attempts to make efficacious
XX CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
XX CC Multivalent vaccines have also been tried but none have successfully
XX CC overcome antigenic variability. The provision of further, complete
XX CC sequences may provide an opportunity to identify secreted or surface
XX CC exposed proteins that may be presumed targets for the immune system and
XX CC which are not antigenically variable or at least more conserved than
XX CC other more variable regions.
XX
XX SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

Alignment Scores:
Pred. No.: 1.21e+06 Length: 1437668
Score: 46.00 Matches: 11
Percent Similarity: 65.38% Conservative: 6
Best Local Similarity: 42.31% Mismatches: 9
Query Match: 35.66% Indels: 0
DB: 21 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AAA81490 (1-1437668)

QY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
Db 411891 GCTTTGGAGCGGTTTGAAAACTCGAAGCGCGCAAAATCCCGAAGCGGCGCATTTG 411950
QY 23 LysAlaAspAsnAlaAla 28
Db 411951 GAAGCGGCTGAAGCGCT 411968

RESULT 34
ABA74678
ID ABA74678 standard; DNA; 234 BP.
XX
XX AC ABA74678;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #22983.
XX
```

CC AAA81259 and AAA81304 to AAA81371 represent PCR primers used in the  
CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to  
CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
CC sequences, which are all used in the exemplification of the present  
CC invention. The nucleic acid sequences, protein sequences, and antibodies  
CC against them, can be used in the manufacture of a composition. The  
CC composition can be used as a medicament (or in the manufacture of a  
CC medicament) for treating, preventing or diagnosing infection due to  
CC *Neisseria meningitidis*. For example, some of the identified proteins could  
CC be components of vaccines against *Neisseria meningitidis* B; against all serotypes;  
CC and/or against all pathogenic *Neisseria meningitidis* strains;  
CC from the bacterium will also facilitate production of biological probes,  
CC particularly organism-specific probes. Attempts to make efficacious  
CC *Neisseria meningitidis* B vaccines have failed mainly due to antigen tolerance.  
CC Multivalent vaccines have also been tried but none have successfully  
CC overcome antigenic variability. The provision of further, complete  
CC sequences may provide an opportunity to identify secreted or surface  
CC exposed proteins that may be presumed targets for the immune system and  
CC which are not antigenically variable or at least more conserved than  
CC other more variable regions.

XX Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;

Alignment Scores:  
Pred. No.: 3.27e+04 Length: 69936  
Score: 46.00 Matches: 11  
Percent Similarity: 65.38% Conservative: 6  
Best Local Similarity: 42.31% Mismatches: 9  
Query Match: 35.66% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAA81479 (1-59936)

QY 3 AlaleuGlualaleuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 22  
AAF21607  
Db 69296 GCTTTGGAAGCGCTTTGGAAACTCGAAGCCGCGCAAAATCCGACGGCAGCATTTG 69237  
QY 23 LysAlaAspAsnAlaAla 28  
Db 69236 GAAGCGGCTGAAGCGGCT 69219

RESULT 31  
AAF21607  
ID AAF21607 standard; DNA: 349980 BP.  
AC AAF21607;  
XX  
XX 13-MAR-2001 (first entry)  
DT  
XX *Neisseria meningitidis* B nucleotide sequence SEQ ID NO:108.  
DE  
XX  
XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; immunogenic; vaccine;  
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
KW ds.  
XX  
XX *Neisseria meningitidis*.  
OS  
XX WO200066791-A1.  
PN  
XX  
XX 09-NOV-2000.  
PD  
XX  
XX 08-MAR-2000; 2000WO-0505928.  
PF  
XX  
XX 30-APR-1999; 99US-0132068.  
PR  
XX 08-OCT-1999; 99WO-0523573.  
PR  
XX 28-FEB-2000; 2000GB-0004695.  
PR  
XX (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
PA  
XX  
XX Pizzia M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V;  
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappuoli R;  
PI Frazer CM, Grandi G;

XX WPI: 2000-647603/62.  
DR  
XX *Neisseria meningitidis* B full length genome sequence and open reading  
PT frames are used to detect, treat and prevent *Neisseria meningitidis* -  
XX  
XX Claim 7; Appendix A; 692pp; English.

CC The present invention describes the full length genome of  
CC *Neisseria meningitidis* B (NMB). The sequences in AAF21544 and AAF21607  
CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
CC sequence was too long to go in a record on its own it was split into 8  
CC sequences which overlap each other at the beginning and end of each  
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
CC *Neisseria* proteins given in AAF21589 to AAF21593, and AAF21589 to  
CC AAF21606 represent PCR primers which are used in the exemplification of  
CC the present invention. The NMB genome and fragments from it have  
CC antibacterial activity, and can be used in vaccines and gene therapy.  
CC *Neisseria* nucleic acids, proteins and/or antibodies which binds to the  
CC proteins can be used in compositions for treating or preventing infection  
CC due to *Neisseria meningitidis* B or as a diagnostic reagent for detecting the  
CC presence of *Neisseria meningitidis* B or of antibodies raised to *Neisseria*  
CC bacteria. Computers, computer memory, computer storage medium or computer  
CC databases can be used in a search to identify open reading frames (ORFs)  
CC or coding sequences within the NMB genome. The DNA sequences provide  
CC further opportunities to find antigenic or immunogenic proteins which are  
CC more effective in vaccines than the outer membrane proteins currently  
CC used.

XX Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

Alignment Scores:  
Pred. No.: 2.41e+05 Length: 349980  
Score: 46.00 Matches: 11  
Percent Similarity: 65.38% Conservative: 6  
Best Local Similarity: 42.31% Mismatches: 9  
Query Match: 35.66% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAF21607 (1-349980)

QY 3 AlaleuGlualaleuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 22  
AAF21607  
Db 111891 GCTTTGGAAGCGCTTTGGAAACTCGAAGCCGCGCAAAATCCGACGGCAGCATTTG 111950  
QY 23 LysAlaAspAsnAlaAla 28  
Db 111951 GAAGCGGCTGAAGCGGCT 111968

RESULT 32  
AAF21607  
ID AAF21607 standard; DNA: 349980 BP.  
AC AAF21607;  
XX  
XX 26-SEP-2001 (first entry)  
DT  
XX  
XX C glutamicum coding sequence fragment SEQ ID NO: 7060.

DE  
XX  
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
KW Corynebacterium glutamicum.  
OS  
XX EP1108790-A2.  
PN  
XX  
XX 20-JUN-2001.  
PD  
XX  
XX 18-DEC-2000; 2000EP-0127688.  
PF  
XX  
XX 16-DEC-1999; 99JP-0377484.  
PR

QY 24 AlaAspAsnAla 27  
 Db 576 TTCGAGACGCT 565  
 RESULT 29  
 ID AAV74485/c  
 XX AAV74485 standard; DNA; 4549 BP.  
 AC AAV74485;  
 XX  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #174.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 XX toxic shock syndrome; ds.  
 XX Staphylococcus aureus.  
 XX  
 XX Key Location/Qualifiers  
 FT misc\_feature 121..180  
 FT /tag= a  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence".  
 FT  
 FT misc\_feature 1921..1980  
 FT /tag= b  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence".  
 FT  
 FT misc\_feature 3721..3780  
 FT /tag= c  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence".  
 FT  
 FT  
 PN EP786519-A2.  
 XX  
 XX 30-JUL-1997.  
 XX  
 XX 07-JAN-1997; 97EP-0100117.  
 XX  
 XX 05-JAN-1996; 96US-0009861.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
 XX Rosen CA;  
 XX  
 XX WPI; 1997-374922/35.  
 XX  
 XX Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 XX stored on computer readable medium and used in the production of  
 XX anti-S.aureus vaccines  
 XX  
 XX Claim 1; Page 849-852; 3271pp; English.  
 XX  
 XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 XX of the invention. The DNA sequences are recorded on a computer readable  
 XX medium, preferably selected from a floppy or hard disk, random access  
 XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 XX the S.aureus DNA sequences allows putative functions to be assigned so  
 XX that protein-encoding or regulatory regions of commercial, therapeutic or  
 XX industrial importance can be obtained. Specifically, sequences which are  
 XX likely to encode antigens have been identified and these polypeptides can  
 XX be used in a vaccine composition against S.aureus infection. The  
 XX polypeptides can also be used in a kit for the immunodetection of

CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 XX  
 XX Sequence 4549 BP; 1387 A; 858 C; 698 G; 1421 T; 185 other;  
 SQ  
 Alignment Scores:  
 Pred. NO.: 1.03e+03 Length: 4549  
 Score: 46.00 Matches: 12  
 Percent Similarity: 60.71% Conservative: 5  
 Best Local Similarity: 42.86% Mismatches: 7  
 Query Match: 35.66% Indels: 4  
 DB: 18 Gaps: 1  
 US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAV74485 (1-4549)  
 QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20  
 Db 2765 TCAGATATATTGGAGCGGTA-----GACATCTTGCACAAATGAAAAAGTAGGCC 2718  
 QY 21 ValValIysAlaAspAsnAlaAla 28  
 Db 2717 ATCGGATATATATATATGTTGCA 2694  
 RESULT 30  
 ID AAA81479/c  
 XX AAA81479 standard; DNA; 69936 BP.  
 AC AAA81479;  
 XX  
 XX 04-DEC-2000 (first entry)  
 XX  
 XX N. meningitidis partial DNA sequence gnm\_27 SEQ ID NO:27.  
 XX  
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 XX antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 XX Meningococcus B; MenB; ds.  
 XX  
 XX Neisseria meningitidis.  
 XX  
 XX WO2000022430-A2.  
 XX  
 XX 20-APR-2000.  
 XX  
 XX 08-OCT-1999; 99WO-US23573.  
 XX  
 XX 09-OCT-1998; 98US-0103794.  
 XX  
 XX 30-APR-1999; 99US-0132068.  
 XX  
 XX (CHTR ) CHIRON CORP.  
 XX  
 XX Frazer CM, Hickey B, Peterson J, Tettelin H, Venter JC;  
 XX Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 XX Rappuoli R, Pizza M;  
 XX  
 XX WPI; 2000-318079/27.  
 XX  
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
 XX used in the diagnosis and treatment of N. meningitidis infection and  
 XX other Neisserial infections, for example, N.gonorrhoea -  
 XX  
 XX Claim 7; Page 547-567; 1760pp; English.  
 XX  
 XX The present invention describes methods of obtaining immunogenic  
 XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 XX represent specifically claimed Neisseria meningitidis genomic DNA  
 XX sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to



```
XX 15-JAN-2002 (first entry)
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX US6294328-B1.
PN
XX
XX 25-SEP-2001.
PD
XX
XX 24-JUN-1998; 98US-0103840.
PF
XX
XX 24-JUN-1998; 98US-0103840.
PR
XX
XX (GENO-) INST GENOMIC RES.
PA
XX
XX Fleischmann RD, White OR, Fraser CM, Venter JC;
PI
XX
XX WPI; 2001-647261/74.
DR
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ
XX
XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
PS
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPRO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
XX Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 2,42e+06 Length: 4411529
Score: 47.00 Matches: 9
Percent Similarity: 54.55% Conservative: 3
Best Local Similarity: 40.91% Mismatches: 10
Query Match: 36.43% Indels: 0
DB: 22 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AAI99682 (1-4411529)
QY 7 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallysAlaAspAsn 26
DB 4044185 TTGATGGACGGCGCGACCTGCTGGAGCGGACAGGTAACCATGCTCGATATCGACAAAC 4044126
QY 27 AlaAla 28
DB 4044125 GGTGCT 4044120
RESULT 26
AAFL1343
XX AAF11343 standard; cDNA; 636 BP.
XX
XX AAF11343;
XX
XX 13-MAR-2001 (first entry)
XX
```

```
DE Aspergillus niger EST SEQ ID NO:3866.
XX
XX Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
XX Aspergillus niger.
OS
XX
XX WO200056762-A2.
PN
XX
XX 28-SEP-2000.
PD
XX
XX 22-MAR-2000; 2000WO-US07781.
PF
XX
XX 22-MAR-1999; 99US-0273623.
PR
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.
PA
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
PI
XX
XX WPI; 2000-594572/56.
DR
XX
XX Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags
XX
XX Claim 87; Page 1728; 3161pp; English.
PS
XX
XX The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 636 BP; 169 A; 155 C; 161 G; 149 T; 2 other;
Alignment Scores:
Pred. No.: 86 Length: 636
Score: 46.00 Matches: 11
Percent Similarity: 64.00% Conservative: 5
Best Local Similarity: 44.00% Mismatches: 9
Query Match: 35.66% Indels: 0
DB: 21 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AAF11343 (1-636)
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 20
DB 149 TCTAAAGCTGCAGAGCCCTTGCCGAGAGAGATCGCGCCGCAAGTAATATGTCAGTCAGCT 208
QY 21 ValValLysAlaAsp 25
```

```
Db 825 GCTGCTGATGGCTTGCA 808
|||||
RESULT 23
ABL25848/c
ID ABL25848 standard; DNA; 4229 BP.
XX AC ABL25848;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 29017.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Claim 1: SEQ ID NO 29017; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 4229 BP; 1287 A; 1022 C; 928 G; 992 T; 0 other;

Alignment Scores:
Pred. No.: 431 Length: 4229
Score: 48.00 Matches: 12
Percent Similarity: 50.00% Conservative: 1
Best Local Similarity: 46.15% Mismatches: 13
Query Match: 37.21% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL25848 (1-4229)
QY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
|||||
Db 1957 GCGGCTGATGCTGGTGGATGCTGATGCTGGTGGATGCTGCTGCTGCT 1898
QY 23 LysAlaAspAsnAlaAla 28
|||||
Db 1897 GCTGCTGATGCTTGCA 1880
RESULT 24
AAI99683/c
AAI99683;
15-JAN-2002 (first entry)
Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
variation; epidemiology; patient treatment; epidemic monitoring; ds.
Mycobacterium tuberculosis.
US6294328-B1.
25-SEP-2001.
980S-0103840.
980S-0103840.
(GENO-) INST GENOMIC RES.
Fleischmann RD, White OR, Fraser CM, Venter JC;
WPI; 2001-647261/74.
Evaluating strain variation of Mycobacterium tuberculosis, comprises
determining the nucleotide sequence of the strain at positions in the
genome corresponding to positions where M. tuberculosis strains CDC
1551 and H37Rv differ
Claim 4: SEQ ID NO 2; 3pp + Sequence Listing; English.
The invention relates to evaluating strain variation within and between
different populations of the tuberculosis bacterium pathogen,
Mycobacterium tuberculosis or related Mycobacterium by determining the
nucleotide sequence of the first strain at positions in the complete
sequence of the genome that correspond to positions that differ in the
nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
H37Rv (AAI99682). The method is useful for evaluating strain variation of
M. tuberculosis and has valuable application in the fields of
tuberculosis genetics, epidemiology, patient treatment and epidemic
monitoring.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from USPTO
at seqdata.uspto.gov/sequence.html?docID=6294328B1.
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Alignment Scores:
Pred. No.: 2,42e+06 Length: 4403765
Score: 47.00 Matches: 9
Percent Similarity: 54.55% Conservative: 3
Best Local Similarity: 40.91% Mismatches: 10
Query Match: 36.43% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AAI99683 (1-4403765)
QY 7 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValValAlaAspAsn 26
|||||
Db 4036256 TTGATGGAGCCGCCGCCCTGCTGGAGGCGGAACAGGTACCATCGTATCGACAC 4036197
QY 27 AlaAla 28
|||||
Db 4036196 GGTGCT 4036191
RESULT 25
AAI99682/c
ID AAI99682 standard; DNA; 4411529 BP.
XX AC AAI99682;
```

/clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II".  
 /note="vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO<sub>2</sub> and HS medium bubbled with 5% CO<sub>2</sub>. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into Lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exassist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 153 a 161 c 206 g 130 t

ORIGIN

Alignment Scores:

Pred. No.: 166 Length: 650  
 Score: 51.00 Matches: 11  
 Percent Similarity: 62.96% Conservative: 6  
 Best Local Similarity: 40.74% Mismatches: 10  
 Query Match: 39.53% Indels: 0  
 DB: 10 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BE337641 (1-650)

Qy 2 AspalateuGluAlaLeuAlaAspGlnThrAspAlateuGlnSerGluGluAlaAlaVal 21  
 ::: :::: ||| ||:::||||| ||:::||||| ||:::|||||

Db 507 GAGAGATCGAGCTGCTGGTGAAGACCGACGACCTGCGCAACCGAGCGGAGCAGTT 566

Qy 22 VallysAlaAspAsnAla 28

||| |||||

Db 567 CCRAAGAGCGGACCGCAGCT 587

RESULT 34

BI633216/c

LOCUS

DEFINITION SD27244.Sprime SD Drosophila melanogaster Schneider L2 cell culture  
 pOT2 Drosophila melanogaster cDNA clone SD27244 5 similar to  
 CGT289: FBan0007289 GO:[ ] located on: 2L 22B8-22B8:: 05/23/2001,  
 mRNA sequence.

ACCESSION

BI633216

VERSION BI633216.1 GI:15535426

KEYWORDS

EST.

SOURCE

ORGANISM

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 660)

Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,

Lewis,S. and Rubin,G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic AE003584: arm:2L [1824960,2149443]

estimated-cyto:22B4-22D2: 05/23/2001

plate: SD.272 row: D column: 8

High quality sequence stop: 529.

Location/Qualifiers

1..660

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="SD27244"

/clone\_lib="SD Drosophila melanogaster Schneider L2 cell

culture pOT2"

FEATURES

source

1..660

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="SD27244"

/clone\_lib="SD Drosophila melanogaster Schneider L2 cell

culture pOT2"

Alignment Scores:

Pred. No.: 172 Length: 666  
 Score: 51.00 Matches: 11  
 Percent Similarity: 53.85% Conservative: 3  
 Best Local Similarity: 42.31% Mismatches: 12  
 Query Match: 39.53% Indels: 0

/lab\_host="DH5-alpha"  
 /note="Vector: pOT2; Site\_1: EcoRI; Site\_2: XhoI; Sized  
 fractionated cDNAs were directly ligated into pOT2.  
 Plasmid cDNA library." 163 t 2 others

BASE COUNT 148 a 189 c 158 g 163 t

ORIGIN

Alignment Scores:

Pred. No.: 170 Length: 660  
 Score: 51.00 Matches: 11  
 Percent Similarity: 53.85% Conservative: 3  
 Best Local Similarity: 42.31% Mismatches: 12  
 Query Match: 39.53% Indels: 0  
 DB: 10 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BI633216 (1-660)

Qy 2 AspalateuGluAlaLeuAlaAspGlnThrAspAlateuGlnSerGluGluAlaAlaVal 21  
 ||| :::: ||| ||||| ||| ||||| |||

Db 226 GATCGCGGAATCATATAGAACACACACAGCAGCAGGAGGAGGAACAAGCGCG 167

Qy 22 VallysAlaAspAsnAla 27

|||::: ||| |||

Db 166 GTACGAGGAGATCGTCC 149

RESULT 35

AI062753/c

LOCUS

DEFINITION GH02073.Sprime GH Drosophila melanogaster head pOT2 Drosophila  
 melanogaster cDNA clone GH02073 Spprime, mRNA sequence.

ACCESSION

AI062753

VERSION AI062753.1 GI:3338592

KEYWORDS

EST.

SOURCE

ORGANISM

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 666)

Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,

Lewis,S. and Rubin,G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic sequence DS06378

Plate: 20 row: G column: 1

High quality sequence stop: 566.

Location/Qualifiers

1..666

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="GH02073"

/clone\_lib="GH Drosophila melanogaster head pOT2"

/sex="male and female"

/dev\_stage="adult"

/lab\_host="DH5 - alpha"

/note="Organ: head; Vector: pOT2; Site\_1: EcoRI; Site\_2:

XhoI; Sized fractionated cDNAs were directly ligated into

pOT2. Plasmid cDNA library."

BASE COUNT 153 a 192 c 157 g 164 t

ORIGIN

Alignment Scores:

Pred. No.: 172 Length: 666  
 Score: 51.00 Matches: 11  
 Percent Similarity: 53.85% Conservative: 3  
 Best Local Similarity: 42.31% Mismatches: 12  
 Query Match: 39.53% Indels: 0

```

VERSION      AI062828.1  GI:3338667
KEYWORDS     EST,
SOURCE       fruit fly,
ORGANISM     Drosophila melanogaster
REFERENCE    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
              1 (bases 1 to 633)
              Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
              Lewis,S. and Rubin,G.M.
TITLE        BDGP/HMI Drosophila EST Project
JOURNAL      Unpublished (2001)
COMMENT      Contact: Stapleton, M.

BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic sequence DS06378
Plate: 21 row: G column: 1
High quality sequence stop: 567.

FEATURES             source
    source
    1..633
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone="CH02173"
    /clone_lib="GH Drosophila melanogaster head pot2"
    /sex="male and female"
    /dev_stage="adult"
    /lab_host="DH5 - alpha"
    /note="Organ: head; Vector: pot2; Site_1: ECORI; Site_2:
    XhoI; Sized fractionated cDNAs were directly ligated into
    pot2. Plasmid cDNA library."
BASE COUNT      140 a 182 c 153 g 158 t
ORIGIN
|||
Alignment Scores:
Pred. No.:      160      Length:      633
Score:          51.00    Matches:      11
Percent Similarity: 53.8%  Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match:     39.53%  Indels:      0
DB:              9      Gaps:         0

US-09-847-539a-6_COPY_59_86 (1-28) x AI062828 (1-633)
QY  2  AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
|||
Db  227  GATCGCGGAATCATATAGAACACACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 168
|||
QY  22  ValLysAlaAspAsnAla 27
|||
Db  167  GTACGAGGAGATCGTGCC 150
|||

RESULT 32
AA391497/c
LOCUS      AA391497/c
DEFINITION LD10172.5prime LD Drosophila melanogaster embryo Bluescript
Drosophila melanogaster cDNA clone LD10172 5prime, mRNA sequence.
ACCESSION  AA391497
VERSION     AA391497.1  GI:2044472
KEYWORDS   EST,
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 643)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.

AI062828.1  GI:3338667
EST,
fruit fly,
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 633)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.

```

```

BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 101 row: F column: 12
High quality sequence stop: 463.

FEATURES             source
    source
    1..643
    /organism="Drosophila melanogaster"
    /db_xref="BDGP_EST:BDcln009443"
    /db_xref="taxon:7227"
    /clone="LD10172"
    /clone_lib="LD Drosophila melanogaster embryo Bluescript"
    /sex="male and female"
    /dev_stage="0 to 24 hours mixed stage embryonic"
    /lab_host="SOLR"
    /note="Organ: embryo; Vector: Bluescript SK; Site_1: ECORI
    ; Site_2: XhoI; Constructed using Stratagene ZAP-CDNA
    Synthesis kit. Oligo dt-primed and directionally cloned at
    EcoRI and XhoI in Bluescript SK(+/-)"
BASE COUNT      140 a 185 c 154 g 164 t
ORIGIN
|||
Alignment Scores:
Pred. No.:      163      Length:      643
Score:          51.00    Matches:      11
Percent Similarity: 53.85%  Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match:     39.53%  Indels:      0
DB:              9      Gaps:         0

US-09-847-539a-6_COPY_59_86 (1-28) x AA391497 (1-643)
QY  2  AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
|||
Db  213  GATCGCGGAATCATATAGAACACACACAGAGGAGGAGGAGGAGGAGGAGGAGG 154
|||
QY  22  ValLysAlaAspAsnAla 27
|||
Db  153  GTACGAGGAGATCGTGCC 136
|||

RESULT 33
BE337641
LOCUS      BE337641
DEFINITION 894047D11.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION  BE337641
VERSION     BE337641.1  GI:9210726
KEYWORDS   EST,
SOURCE     Chlamydomonas reinhardtii.
ORGANISM   Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
1 (bases 1 to 650)
Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 2
Unpublished (2000)
Contact: Elizabeth H. Harris
DCMB Box 91000
Duke University
Durham, NC 27708-1000, USA
Tel: 919 613 8164
Fax: 919 613 8177
Email: chlamy@duke.edu.
Location/Qualifiers
1..650
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"

```











```

REFERENCE
AUTHORS      1 (bases 1 to 558)
              Kohara,Y., Shin-I.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
              and Sugano,S.
TITLE        A complementary view of the C.elegans genome
JOURNAL      Unpublished (2001)
COMMENT      Contact: Yuji Kohara
              National Institute of Genetics
              Genome Biology Lab.
              Yata 1111, Mishima, Shizuoka 411, Japan
              Tel: 81-559-81-6854
              Fax: 81-559-81-6855
              Email: ykohara@lab.nig.ac.jp.
FEATURES
source       Location/Qualifiers
              1..558
              /organism="Caenorhabditis elegans"
              /strain="N2"
              /db_xref="taxon:6239"
              /clone="YK858ell"
              /clone_lib="unpublished oligo-capped cDNA library, stage
              L1"
              /sex="Hermaphrodite"
              /tissue_type="whole animal"
              /dev_stage="L1"
              164 a 146 c 142 g 105 t 1 others
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 134 Length: 558
Score: 51.00 Matches: 9
Percent Similarity: 68.00% Conservative: 8
Best Local Similarity: 36.00% Mismatches: 8
Query Match: 39.53% Indels: 0
DB: Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AU206045 (1-558)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 236 TCAGATGCTATTCAAGAGCTTTCCGATCAATCGACCACTTCAAAAGCAAAAGGGAAGA 295

Qy 21 ValVallyAlaAsp 25
::: ||| ::
Db 296 ATTCAGAAGGAGGAA 310

RESULT 21
BF486089/c 563 bp mRNA linear EST 19-APR-2001
LOCUS
DEFINITION
BF486089 AT20267.5:prime AT Drosophila melanogaster adult testes pOTB7
Drosophila melanogaster cDNA clone AT20267 5 similar to CG7289:
FBan0007289 located on: 2L 22B8-22B8:: 04/09/2001, mRNA sequence.
ACCESSION
VERSION BF486089.2 GI:13697978
KEYWORDS
SOURCE EST.
ORGANISM Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS 1 (bases 1 to 563)
Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman
,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan
,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P.,
Liao,G., Miranda,A., Mista,S., Mungall,C.J., Nunoo,J., Pacleb,J.,
Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E.,
Celniker,S. and Rubin,G.M.
BGPH/HMI AT Drosophila EST Project
Unpublished (2000)
On Dec 6, 2000 this sequence version replaced gi:11569390.
Contact: Stapleton, M.
BGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798

REFERENCE
AUTHORS      1 (bases 1 to 567)
              Lawrence Berkeley National Lab
              One Cyclotron Rd, Berkeley, CA 94720, USA
              Fax: 510 486 6798
              Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
              hit genomic sequence DS06378
              Plate: 30 row: C column: 6
              High quality sequence stop: 454.
FEATURES
source       Location/Qualifiers
              1..567
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"

```

```

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003584: arm:2L [1824960,2149443]
estimated-cyto:22B4-22D2: 04/09/2001
Plate: AT.202 row: F column: 7
High quality sequence stop: 562.
FEATURES
source       Location/Qualifiers
              1..563
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"
              /clone="AT20267"
              /clone_lib="AT Drosophila melanogaster adult testes pOTB7"
              /sex="male"
              /dev_stage="0-3 day old Ore-R males"
              /lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
              AT.121-AT.319: DH5-alpha Tona"
              /note="Organ: ADULT testes; Vector: pOTB7; Site:1: EORI;
              Site_2: XhoI; The mRNA for the testis library was made
              from testes and seminal vesicles hand dissected from 0-3
              day old Ore-R males. RNA kindly provided by the lab of
              Margaret Fuller. Sized fractionated cDNAs were directly
              ligated into pOTB7. Plasmid cDNA library."
              126 a 159 c 141 g 137 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 136 Length: 563
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
DB: Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x BF486089 (1-563)

Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 228 GATCGCGGAATCATATAGAGACCACACAGCAGGAGGAGGAGGAGGAGGAGGAGG 169

Qy 22 VallyAlaAspAsnAla 27
||||: ||| |||
Db 168 GTACGAGGAGATCGTCC 151

RESULT 22
AI063297/c 567 bp mRNA linear EST 19-APR-2001
LOCUS
DEFINITION
AI063297 GH03030.5:prime GH Drosophila melanogaster head pOT2 Drosophila
melanogaster cDNA clone GH03030 5:prime, mRNA sequence.
ACCESSION
VERSION AI063297.1 GI:3339241
KEYWORDS
SOURCE EST.
ORGANISM Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS 1 (bases 1 to 567)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BGPH/HMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic sequence DS06378
Plate: 30 row: C column: 6
High quality sequence stop: 454.
FEATURES
source       Location/Qualifiers
              1..567
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"

```











```

KEYWORDS
SOURCE Saccharomyces kluyveri.
ORGANISM Saccharomyces kluyveri
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 1163)
Neuveglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
Galliardin,C. and Casaregola,S.
Genomic exploration of the hemiascomycetous yeasts: 9.
JOURNAL Saccharomyces kluyveri
MEDLINE FEBS Lett. 487 (1), 56-60 (2000)
REFERENCE
2 (bases 1 to 1163)
Souclet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpartuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL Saccharomyces kluyveri
MEDLINE FEBS Lett. 487 (1), 3-12 (2000)
REFERENCE
3 (bases 1 to 1163)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
Location/Qualifiers
1. .1163
/organism="Saccharomyces kluyveri"
/strain="CBS 3082"
/db_xref="taxon:4934"
/clone="AU0AA012D04"
/clone_lib="AUGAA"
/note="end : T3"
BASE COUNT 359 a 212 c 291 g 297 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 259 Length: 1163
Score: 52.00 Matches: 12
Percent Similarity: 71.43% Conservative: 3
Best Local Similarity: 57.14% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 12 Gaps: 0
US-09-847-539a-6_COPY_59_86 (1-28) x CNS06NI7 (1-1163)
QY 5 GluaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValLysAla 24
|||||:||||| ||||| |||||||||:|||||
DB 398 GAGCTGTACTGCTGTGAGTGCAGTGCAGATTCAGAGAGCCCTCGTGGAGATGCA 457
QY 25 Asp 25
DB 458 GAG 460
RESULT 10
BF496493/c 243 bp mRNA linear EST 19-APR-2001
LOCUS AT10425.5prime AT Drosophila melanogaster adult testes pOTB7
DEFINITION Drosophila melanogaster cDNA clone AT10425 5 similar to CG7289:
KEYWORDS fruit fly.
SOURCE EST.
ACCESSION FBAN0007289 located on: 2L 22B8-22B8; : 04/08/2001, mRNA sequence.
BF496493
VERSION BF496493.2 GI:13694046
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 243)
Stapleton,M., Brokstein,P., Hong,L., Aqbayani,A., Baxter,E., Berman
,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan
,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P.,
Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J.,
Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E.,
Celniker,S. and Rubin,G.M.
BDGP/HHMI AT Drosophila EST Project
Unpublished (2000)
JOURNAL
COMMENT On Dec 6, 2000 this sequence version replaced gi:11579867.
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: AT.104 row: C column: 1
High quality sequence stop: 235.
FEATURES
source
Location/Qualifiers
1. .243
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="AT10425"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
AT.121-AT.319: DH5-alpha Tona"
/note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."
BASE COUNT 53 a 65 c 56 g 68 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 41.7 Length: 243
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
DB: 10 Gaps: 0
US-09-847-539a-6_COPY_59_86 (1-28) x BF496493 (1-243)
QY 2 AspaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
|||||:||||| ||||| |||||||:|||||
DB 228 GTACGGCGGAATCATATAGAAAGACACACAGCAGGAGGAGGAGGAGGAGCGG 169
QY 22 VallysAlaAspaSnAla 27
|||||: ||| |||
DB 168 GTACGAGGAGATCGTGCC 151
RESULT 11
AA140863 333 bp mRNA linear EST 29-NOV-1998
LOCUS CK00525.5prime CK Drosophila melanogaster embryo BlueScript
DEFINITION Drosophila melanogaster cDNA clone CK00525 5prime, mRNA sequence.
ACCESSION AA140863
VERSION AA140863.1 GI:1704336
KEYWORDS EST.
SOURCE fruit fly.

```

```

/clone="BACR32D16"
/Note="end : T7"
BASE COUNT 268 a 329 c 226 g 231 t 47 others
ORIGIN

Alignment Scores:
Pred. No.: 166 Length: 1101
Score: 53.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.09% Indels: 0
DB: 12 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x CNS00FX5 (1-1101)

QY 5 GluAlaLeuAlaLeuAlaLeuGlnSerGluGluAlaAlaValLysAla 24
||| |||||: ||||| :|||: ||||| |||
Db 263 GAGGAGGAGCTGACGAGGAGCGCCGCGAGGAGAGGATGCCGCGCCGCC 204

QY 25 AspAsnAlaAla 28
||| |||||
Db 203 GATGCTGCGCG 192

RESULT 7
AA941632/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AA941632 AA941632 438 bp mRNA linear EST 19-APR-2001
LD25873.5prime LD Drosophila melanogaster embryo pot2 Drosophila
melanogaster cDNA clone LD25873 5prime, mRNA sequence.

AA941632.1 GI:3101545
EST.
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Drosophila melanogaster
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 438)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/RHMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 258 row: G column: 1
High quality sequence stop: 226.

FEATURES
Source
Location/Qualifiers
1..438
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD25873"
/clone_lib="LD Drosophila melanogaster embryo pot2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XL1 Blue"
/Note="Organ: embryo; Vector: pot2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2."
BASE COUNT 98 a 125 c 99 g 116 t
ORIGIN

Alignment Scores:
Pred. No.: 65.9 Length: 438
Score: 52.00 Matches: 11
Percent Similarity: 55.56% Conservative: 4
Best Local Similarity: 40.74% Mismatches: 12
Query Match: 40.31% Indels: 0
DB: 9 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AA941632 (1-438)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
::: ||| ||||| ||| ||||| |||
Db 231 AACGATCGCGGAATCATATAGAGACCACAGACGAGGAGGAGGAGCAAGCG 172

QY 21 ValValLysAlaAspAsnAla 27
|||: ||| |||
Db 171 GCGGTACGAGGAGATCGTGCC 151

RESULT 8
BE056346/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BE056346 BE056346 553 bp mRNA linear EST 08-JUN-2000
00257 leafy spurge Lambda HybridAP 2.1 two-hybrid vector cDNA
Library Euphorbia esula cDNA clone 25D 5' similar to Arabidopsis
Unknown Protein (accession# AAF24521), mRNA sequence.

BE056346.1 GI:8383403
EST.
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

leafy spurge.
Euphorbia esula
Euphorbia
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Malpighiales; Euphorbiaceae; Euphorbia.
1 (bases 1 to 553)
Anderson,J.V. and Horvath,D.P.
Identification of mRNAs expressed in underground adventitious buds
of Euphorbia esula (leafy spurge)
Unpublished (2000)
Contact: Anderson JV
Plants Science Research
USDA/ARS, Biosciences Research Lab
1605 Albrecht Blvd., PO Box 5674, Fargo, ND 58105, USA
Tel: 701 239 1263
Fax: 701 239 1252
Email: andersjv@fargo.ars.usda.gov
Seq primer: PAD5.

FEATURES
Source
Location/Qualifiers
1..553
/organism="Euphorbia esula"
/db_xref="taxon:3993"
/clone="25D"
/clone_lib="leafy spurge Lambda HybridAP 2.1 two-hybrid
vector cDNA Library"
/tissue_type="underground adventitious buds"
/dev_stage="3-day induced (decapitated)"
BASE COUNT 161 a 117 c 116 g 159 t
ORIGIN

Alignment Scores:
Pred. No.: 91.3 Length: 553
Score: 52.00 Matches: 12
Percent Similarity: 65.38% Conservative: 5
Best Local Similarity: 46.15% Mismatches: 9
Query Match: 40.31% Indels: 0
DB: 9 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x BE056346 (1-553)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
|||: ||||| ||| ||||| ||| |||: ||| |||
Db 541 TCAGAAGCAGCTTTCTTCTGCTGATGCCACAAAGCGATGAGATCAGCAACCTTGGCC 482

QY 21 ValValLysAlaAspAsn 26
::: |||||
Db 481 ATTTCCATACACGATAT 464

RESULT 9
CNS06NI7
LOCUS
DEFINITION
ACCESSION
VERSION

CNS06NI7 1163 bp DNA linear GSS 17-JUN-2001
T3 end of clone AU0AA012D04 of library AU0AA from strain CBS 3082
of Saccharomyces kluyveri, genomic survey sequence.
AL406805
AL406805.1 GI:12171287

```



```

source
1. .788
/organism="Tetrahymena thermophila"
/db_xref="taxon:5911"
/clone_lib="Chilcoat/Turkewitz cDNA (large fraction)"
/note="Vector: Bluescript2 SK+; Details on library
preparation can be found in Chilcoat and Turkewitz (2001)
Proc. Natl. Acad. Sci USA, 98: 8709-8713."
BASE COUNT 248 a 219 c 121 g 199 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 71.5 Length: 788
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 10 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x BM400696 (1-788)

Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
||||| :||||| ||| ||||||||| :||| |||:
Db 265 GATGCTCAACAGCTCTCTAAGACACACTGCTCTCAATGCTGCCGAGAAGCTCTT 324
||||| :||||| ||| ||||||||| :||| |||:
Qy 22 ValLysAlaAsp 25
|||||
Db 325 GCCAATGCTGAA 336
|||||

RESULT 2
BM396025
LOCUS
DEFINITION
5009-0-15-G03.t.2 Chilcoat/Turkewitz cDNA (large fraction)
Tetrahymena thermophila cDNA, mRNA sequence.
ACCESSION
BM396025
VERSION
BM396025.1 GI:18196078
KEYWORDS
EST.
SOURCE
Tetrahymena thermophila.
ORGANISM
Tetrahymena thermophila
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE
1 (bases 1 to 861)
Turkewitz,A.P., Karrer,K.M., Jahn,C., Orias,E., Kirk,K.E., Frankel
,J. and Klobutcher,L.
EST from Tetrahymena thermophila, strain CU428.1, growing cells
Unpublished (2002)
JOURNAL
COMMENT
Contact: Turkewitz AP
Molecular Genetics and Cell Biology
University of Chicago
920 E. 58th Street, Chicago, IL 60637, USA
Tel: 773 702 4374
Fax: 773 702 3172
Email: apturkew@midway.uchicago.edu
Seq primer: T3.
FEATURES
source
1. .861
Location/Qualifiers
/organism="Tetrahymena thermophila"
/db_xref="taxon:5911"
/clone_lib="Chilcoat/Turkewitz cDNA (large fraction)"
/note="Vector: Bluescript2 SK+; Details on library
preparation can be found in Chilcoat and Turkewitz (2001)
Proc. Natl. Acad. Sci USA, 98: 8709-8713."
BASE COUNT 278 a 222 c 151 g 207 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 81 Length: 861
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 10 Gaps: 0

source
1. .788
/organism="Tetrahymena thermophila"
/db_xref="taxon:5911"
/clone_lib="Chilcoat/Turkewitz cDNA (large fraction)"
/note="Vector: Bluescript2 SK+; Details on library
preparation can be found in Chilcoat and Turkewitz (2001)
Proc. Natl. Acad. Sci USA, 98: 8709-8713."
BASE COUNT 248 a 219 c 121 g 199 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 71.5 Length: 788
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 10 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x BM400696 (1-788)

Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
||||| :||||| ||| ||||||||| :||| |||:
Db 265 GATGCTCAACAGCTCTCTAAGACACACTGCTCTCAATGCTGCCGAGAAGCTCTT 324
||||| :||||| ||| ||||||||| :||| |||:
Qy 22 ValLysAlaAsp 25
|||||
Db 325 GCCAATGCTGAA 336
|||||

RESULT 2
BM396025
LOCUS
DEFINITION
5009-0-15-G03.t.2 Chilcoat/Turkewitz cDNA (large fraction)
Tetrahymena thermophila cDNA, mRNA sequence.
ACCESSION
BM396025
VERSION
BM396025.1 GI:18196078
KEYWORDS
EST.
SOURCE
Tetrahymena thermophila.
ORGANISM
Tetrahymena thermophila
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE
1 (bases 1 to 861)
Turkewitz,A.P., Karrer,K.M., Jahn,C., Orias,E., Kirk,K.E., Frankel
,J. and Klobutcher,L.
EST from Tetrahymena thermophila, strain CU428.1, growing cells
Unpublished (2002)
JOURNAL
COMMENT
Contact: Turkewitz AP
Molecular Genetics and Cell Biology
University of Chicago
920 E. 58th Street, Chicago, IL 60637, USA
Tel: 773 702 4374
Fax: 773 702 3172
Email: apturkew@midway.uchicago.edu
Seq primer: T3.
FEATURES
source
1. .861
Location/Qualifiers
/organism="Tetrahymena thermophila"
/db_xref="taxon:5911"
/clone_lib="Chilcoat/Turkewitz cDNA (large fraction)"
/note="Vector: Bluescript2 SK+; Details on library
preparation can be found in Chilcoat and Turkewitz (2001)
Proc. Natl. Acad. Sci USA, 98: 8709-8713."
BASE COUNT 278 a 222 c 151 g 207 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 81 Length: 861
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 10 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x BM396025 (1-861)

Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
||||| :||||| ||| ||||||||| :||| |||:
Db 501 GATGCTCAACAGCTCTCTAAGACACACTGCTCTCAATGCTGCCGAGAAGCTCTT 560
||||| :||||| ||| ||||||||| :||| |||:
Qy 22 ValLysAlaAsp 25
|||||
Db 561 GCCAATGCTGAA 572
|||||

RESULT 3
BJ012116
LOCUS
DEFINITION
BJ012116 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA171D09 5',
mRNA sequence.
ACCESSION
BJ012116
VERSION
BJ012116.1 GI:17354928
KEYWORDS
EST.
SOURCE
Japanese medaka.
ORGANISM
Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE
1 (bases 1 to 613)
Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
JOURNAL
COMMENT
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1. .613
Location/Qualifiers
/organism="Oryzias latipes"
/db_xref="taxon:8090"
/clone="MF01SSA171D09"
/clone_lib="MF01SSA cDNA"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
BASE COUNT 134 a 136 c 239 g 104 t
ORIGIN

Alignment Scores:
Pred. No.: 72.9 Length: 613
Score: 53.00 Matches: 12
Percent Similarity: 60.71% Conservative: 5
Best Local Similarity: 42.86% Mismatches: 11
Query Match: 41.09% Indels: 0
DB: 10 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x BJ012116 (1-613)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 20
||||| :||||| ||| ||||||||| :||| |||:
Db 8 TCACAGCGCGCGGAGGAGCGCGGAGGAGGCGGCTTTAACCGGAGGAGGCGCG 67
||||| :||||| ||| ||||||||| :||| |||:
Qy 21 ValValLysAlaAspAlaAla 28
|||||
Db 68 GCTTTAACCGGAGGAGGCGCGCT 91
|||||

RESULT 4
BJ011078
LOCUS
DEFINITION
BJ011078 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA153F11 5',
mRNA sequence.
ACCESSION
BJ011078

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:12:56 ; Search time 332.257 Seconds  
(without alignments)  
1137.418 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALALADQTDALQSEAAVVKADNAA 28

Scoring table:

BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-O/cgn2\_1/USPTO\_spool/US09847539/runat\_10102002\_092548\_3435/app\_query.fasta\_1.526  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539.ecgn\_1\_1\_763/erunat\_10102002\_092548\_3435 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THRAIDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estopl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	41.9	788	10	BM400696
2	54	41.9	861	10	BM396025

3	53	41.1	613	10	BJ012116
4	53	41.1	615	10	BJ011078
5	53	41.1	624	10	BJ013747
6	53	41.1	1101	12	CNS000FX5
7	52	40.3	438	9	AA941632
8	52	40.3	553	9	BE056346
9	52	40.3	1163	12	CNS06NI7
10	51	39.5	243	10	BF496493
11	51	39.5	333	9	AA140863
12	51	39.5	383	9	AI552971
13	51	39.5	387	9	AI467576
14	51	39.5	454	9	AA423358
15	51	39.5	491	9	AI405338
16	51	39.5	529	9	AU101404
17	51	39.5	535	9	AA540164
18	51	39.5	548	9	AW659168
19	51	39.5	553	10	BM431282
20	51	39.5	558	9	AU206045
21	51	39.5	563	10	BF486089
22	51	39.5	567	9	AI063297
23	51	39.5	567	9	AI388871
24	51	39.5	567	10	BI637415
25	51	39.5	568	9	AI109571
26	51	39.5	569	9	AA392809
27	51	39.5	581	9	AI388559
28	51	39.5	583	9	AI108834
29	51	39.5	615	12	AZ698116
30	51	39.5	622	9	BB201301
31	51	39.5	633	9	AI062828
32	51	39.5	643	9	AA391497
33	51	39.5	650	10	BE337641
34	51	39.5	660	10	BI633216
35	51	39.5	666	9	AI062753
36	51	39.5	666	12	AZ699109
37	51	39.5	710	10	BI637204
38	51	39.5	797	10	BF494445
39	50	38.8	445	10	BG840448
40	50	38.8	483	10	BF487636
41	50	38.8	495	9	AU208069
42	50	38.8	528	9	AV393274
43	50	38.8	547	10	BG051545
44	50	38.8	550	9	AI919727
45	50	38.8	552	9	AU205634

## ALIGNMENTS

RESULT 1  
BM400696

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BM400696 788 bp mRNA linear EST 17-JAN-2002  
5009-0-77-F09.t.1 Chilcoat/Turkewitz cdNA (large fraction)  
Tetrahymena thermophila cdNA, mRNA sequence.

BM400696

EST

Tetrahymena thermophila

Tetrahymena thermophila

Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;

Hymenostomatida; Tetrahymenina; Tetrahymena.

1 (bases 1 to 788)

Turkewitz, A.P., Karrer, K.M., Jahn, C., Orlas, E., Kirk, K.E., Frankel

, J. and Klobutcher, L.

EST from Tetrahymena thermophila, strain CU428.1, growing cells

Unpublished (2002)

Contact: Turkewitz AP

Molecular Genetics and Cell Biology

University of Chicago

920 E. 58th Street, Chicago, IL 60637, USA

Tel: 773 702 4374

Fax: 773 702 3172

Email: apturkew@midway.uchicago.edu

Seq primer: T3.

Location/Qualifiers

FEATURES

```

terminator 10601..10643
/gene="gyra"
10729..12243
/gene="lin0008"
10729..12243
/gene="lin0008"
/note="similar to cardiolipin synthase"
/codon_start=1
/transl_table=11
/protein_id="CAC95241.1"
/db_xref="GI:16412429"
/translation="MRKLVQFLFIAVFLVEVYLINQAAILFLITSGIIQLCCVILIA
IRLLFDORNTSSKVAIVIAVIFILPVLGTSILVFGRRNPATRRFSTAQVMEKAKLINA
IHAIPNTNEKLPRLSRIALHTSIEPIKNGKIEILTNGEETFPVLLDALRKAENHIH
IQYIYFTDAISTEIRDLVEKAKSGVEVRFMDGLSGSKSLAFAPLKEAGVSIHA
FOPTASPIVTRTANLRHRIIVIDGIGFTGGLNIGEEVRSNTPDFRVWRDTHMKIT
GOAVTELQESFLNDWIMENAGAADOQISEAGLQOVFSPVDGDEWAQVITYGGPYDK
EKWVRDSMLDLIDSAKESVMTVSPYFYPDEALAVIRRVAMSGVDVVRVIIPGKDRGI
SFHGSNAYVTMICAGAKMTAYADDSFVHAKMLVDGTRAAIGTANFEDVRSFRUNHEL
MVLYDESEAMHLKRDFFKDFEDSRFLTKMDMENKPLLRITRIKEVLSLLSPIL"
12247..12777
/gene="lin0009"
12247..12252
/gene="lin0009"
12259..12777
/gene="lin0009"
/note="similar to spermidine N1-acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="CAC95242.1"
/db_xref="GI:16412430"
/translation="MSGDLKRLPLEREDLKFVHRLNNDKIMSYWFEEPEAFVELQE
LYDKHIHQDSERFILDGQVGLVELMEIDYIHRRAEFQIIDIPIFGHGYAVSAT
KLAMKYAFHVLNLHKLVLVVDKVKNEKAHVVEKVGFIREGELIDEPFVDGTYHDAIRM

```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
6.37e+03	51.00	231450	10
Percent Similarity:	61.54%	Conservative:	6
Best Local Similarity:	38.46%	Mismatches:	10
Query Match:	39.53%	Indels:	0
DB:	1	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AL596163 (1-231450)

Oy 1 SerAspAlaLeuAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
 : : : : : ||| ||||| : : : : : ||| : : : : :  
 Db 43340 GCAGAAACAAAGACAGATTAGCTGCGGACCCGATATTAATAGATAAGAGATACAGT 43399  
 : : : : : ||| ||||| : : : : : ||| : : : : :  
 Oy 21 ValValLysAlaAspAsn 26  
 : : : : : ||| : : : : :  
 Db 43400 GGAGTAGAGAAAGAAAC 43417

Search completed: October 13, 2002, 03:04:30  
 Job time : 503.262 secs

Tierrez,A., Vazquez-Boland,J.A., Voss,H., Wehland,J. and Cossart,P.  
Comparative genomics of *Listeria species*  
Science 294 (5543), 849-852 (2001)  
21537279  
2 (bases 1 to 231450)  
Glaser,P., Frangeul,L. and Rusniok,C.  
Direct Submission  
Submitted (09-JUL-2001) Glaser P., Institut Pasteur, Genomique des  
Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris  
Cedex 15, FRANCE  
E-mail: pglaser@pasteur.fr  
Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.  
Location/Qualifiers  
1. -231450  
/organism="Listeria innocua"  
/strain="Clp11262"  
/db\_xref="taxon:1642"  
306. .311  
/gene="dnaA"  
306. .1674  
/gene="dnaA"  
319. .1674  
/gene="dnaA"  
/codon\_start=1  
/transl\_table=11  
/product="Chromosomal replication initiation protein DnaA"  
/protein\_id="CAC95234.1"  
/db\_xref="GI:16412422"

COMMENT  
E-mail: pglaser@pasteur.fr  
Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.

FEATURES  
Location/Qualifiers

Source

RBS

gene

CDS

gene

CDS

terminator

gene

CDS

gene

CDS

gene

CDS

gene

CDS

RBS

gene

CDS

/gene="lin0004"  
4645. .4863  
/gene="lin0004"  
/note="similar to B. subtilis yaaA protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC95237.1"  
/db\_xref="GI:16412425"  
/translation="MAETVKINSEPVTLQQLQMLQIDVSTVGMAKALSENTIYINGE  
QNRKGLRNGDVILVPCGVKVIKIQG"

4870. .5982  
/gene="RecF"  
4870. .5982  
/gene="RecF"  
/codon\_start=1  
/transl\_table=11  
/product="RecF protein"  
/protein\_id="CAC95238.1"  
/db\_xref="GI:16412426"

/translation="MHLEISIVLRNERNYENLELEFSPSVNVFLGNAQKTNLLEAVL  
MLAKAKSHRTTNDKDFIMWEKEEAKMEGRVVRKQVTPLELAITQGRKRAKVNHLQEQK  
KLSQVGNLNVVIFAPEDLSLVKGGAPVRRFLNMEIGQMFIYLNHSEYQRILQQR  
NOYLRMLQNRKRVDPMLLDILTEQFADVAINLTKRADFIQLEAYAAPIHHQISRLG  
ETLKIYKASVTLNGDDPEVKADLLQKMEIKQREIDRGVTLIGPHRDSLFYINGQ  
NYQDFSGOQORTTALSVKLAEDILIHETGEVPPVLLDDVLSELDDYRQSHLGAIE  
GKVFQVTTTSGIDHNTLRQATTFYVEKGVKKS"

6031. .7971  
/gene="gyrB"  
6031. .7971  
/gene="gyrB"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase subunit B"  
/protein\_id="CAC95239.1"  
/db\_xref="GI:16412427"

/translation="MSEINITNVHESADYNEDQIQVLEGLAVRRKPGMYIGSTSOR  
GLHLVWEIVDNAIDEALAGFCTEITIEADNSITVRDNGRGIPTGINKETGRPTVE  
VIPTVLHAGGKFGGGYKVGSLHGVASVNALSTSLVYVHREGOKYORFERGDV  
VMDMEVQGETDRTGTVIHTPTDQIFTEETEDFDLTRTRRELAFNLNGLTISIEDK  
REEHKVDFHYEGGTHESGFKTALTRINDYARNKLPKDSDDNLSDGDRVREGLTAI  
NISFANNIHTYEGGTHESGFKTALTRINDYARNKLPKDSDDNLSDGDRVREGLTAI  
ISIKHPDQEGQTKLCNSEARSITDKLFSALNKPMMENPDVAKKIVEKGVVASR  
ARLAARAREVARKSSGLEISLPGKLADCSSRNPEISELYIVEGDSAGSAKQGRDR  
LFOATLPTRGKILNVKARLDRILANEEITRTIETAMCTGFGGDFQVSRVYKLLIMT  
DADVDGAHRTULLTLPTRYMRPLLDAGTIIYAQPLIQIKGKQIEVYVSSQLEDDY  
LASLGDPTKYSIQRYKGLGEMNPEQLWDTTMMPEHRTLLQVNIKDAIDAFETFMLMG  
DVEPRRKFIEDNAYQVKNLDV"

8049. .8054  
/gene="gyrA"  
8049. .10643  
/gene="gyrA"  
8066. .10594  
/gene="gyrA"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase subunit A"  
/protein\_id="CAC95240.1"  
/db\_xref="GI:16412428"

/translation="MAETPNQRITELNKNKERTSFLDYAMSVIVARALPOVRDGLKP  
VHRRILYAMNDLGMTSDKAYKKSAIVIGEVIGKYPHGDVTAVYFTVMRMAQDFSYRNM  
LVDHGNGFSGVDGMDMAAMRYTEARMSKISMLLRDINKDITDYADNYDGSREPVIL  
PARFNNLLVNGSSGTAAGMATNPTPHLGEVIDGVLALSHDPEISTRLMEHILPGPDF  
PTAGMIMGSGIRRAYESGRGSITVRGRVDIEBKNGKETIIVITELPQVKNARLVER  
IAELAREKIDGITSLUNDESRRMRIVIEVRDJSASVIVNNLPMKMTALQTFGINM  
LALVNHKPVNLKLELILYHLQHVQVRRRTFELRKAERALEGLRIALDNIDA  
ITKLKRGKTSADVAKELGTLQNLSDKQAQALDMRLQRLTGLEKREKIEEYQNLVAL  
INDKAILLADDERLLEIREELEEIKVKYAKRRRTTELAGLVLSDEDEDLIPEEVVAI  
TLTKGYTLRPLSLTYRSQRCGREGIQCQGMSTHEDDEHVLVATSHDTLTPTTKGVK  
YRSKGYEPEYCGTAKGPIINLLGIESQBQVNAVNLSEFTDDSYLFTTTHGVKVR  
TTLSQFARILQSGRLAVELRNDLISIVOMTQGSNMIITATKGSITYPEGNRMVG  
PTAAGVGRIRREDDEVIGMEVLEDEKQVLYVTEKGYQKQTPASQYPLNRNGMGVKT  
VTTTEKGNLVANKVTGEEIDLMLTMSGVILRFETDVTVSQTRGSAWGKVLRLDEDE  
KVATVAKVPKEDEVELEEEIDETLITQVPDESFEADAPGSDIEE"



```

/db_xref="GI:17428525"
/translation="MEAVLIVDSRTGTARQVAQLAQSGLAEVGLRPRAGFSG
DARCVLENLNRASYECRALDGEQVIVTAPVWAGRLASPMRIFLADRLPFPVRL
PRCA"
1570..1788
/gene="RSC1510"
/notes="RS03794"
1570..1788
/gene="RSC1510"
/function="miscellaneous; unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD15212.1"
/db_xref="GI:17428526"
/translation="MCVMAARGAFNAIEETAGLTSTLLAPVLALLQDVASGLSQDEI
SGFIHOVRTASDRARRAANLSPSEA"
1839..2207
/gene="RSC1511"
/notes="RS03793"
1839..2207
/gene="RSC1511"
/function="miscellaneous; hypothetical/global homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD15213.1"
/db_xref="GI:17428527"
/translation="MAQLTAAQHQHIALLDQEARLHROLTELESVPADAPPAVEP
YEVDIADLEASERADMRNHYTELAQIVARERLADRGYGLCVDCGEAIFFLRLQ
AQTAQCCVACQKRERRNA"
complement(2254..2532)
/gene="RSC1512"
/notes="RS03792"
complement(2254..2532)
/gene="RSC1512"
/function="miscellaneous; unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD15214.1"
/db_xref="GI:17428528"
/translation="MKSIEPPGPRPOYSATNLTLTCVANVDGSPACVYATAEDHF
GAKSYLOEDJLLALETHRDIAEDMARALFKLTDHNLVIRSGHFRFGM"
complement(2866..3321)
/gene="rnhA"
complement(2866..3321)
/gene="rnhA"
/EC_number="3.1.26.4"
/function="macromolecule metabolism; macromolecule
degradation; degradation of rna"
/notes="Product confidence : probable
Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"

```

```

/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="PROBABLE RIBONUCLEASE HI PROTEIN"
/protein_id="CAD15215.1"
/db_xref="GI:17428529"
/translation="MOEVTIVYSDGACKGNPGI.GGWGTVLVSGGHEKELEFGGAVTTNN
RMELMAVIEAFRALRPCRKYVYTDQYVQKGISEWLAKGKRWKTKADKKPKVKNDDL
WRTLDLVVYTHEVSWHWKGHAGHPGNERADALANKVEIARQAQAGA"
complement(3365..4138)
/gene="RSC1514"
/notes="RS03790"
complement(3365..4138)
/gene="RSC1514"
/function="miscellaneous; hypothetical/partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD15216.1"
/db_xref="GI:17428530"
/translation="MEGFHALQLGHPGALRENRIPLIARVITDDIDPGAPIDTLDPA
DDPGCEADAEPVPOQATPRVICRYDELFPASQSIDLVALPHVLFTDDPHEVLRE
VARVLMPEGRVITGNPLSLMGMRGMRLETSELPASQMIATRLKDWLKLIGF
DIVRGFGCCPPNTRDKWLQRTAFMEKAGDRWWPIFGAVYMLQAVKRVSRVLVGPA
WKRKSPITLVPGATPVATPSSHTTKTPPVAARDKSGA"
4227..5081
/gene="gloB"
/notes="RSC1515; RS03789"
4227..5081
/gene="gloB"

Alignment Scores:
Pred. No.: 5.54e+03 Length: 202050
Score: 51.00 Matches: 11
Percent Similarity: 68.42% Conservatative: 2
Best Local Similarity: 57.89% Mismatches: 6
Query Match: 39.53% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AL646065 (1-202050)

Qy 9 AspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsuAla 27
||||| |||||||: ||| ||||| |||||: |||
Db 3887 GACCAGGTCGATGCTCTCGAGCGCAACGCGAGCTCGTAGCGGAGATCACGCG 3943

RESULT 40
AL596163 231450 bp DNA linear BCT 04-DEC-2001
LOCUS Listeria innocua Clip11262 complete genome, segment 1/12.
DEFINITION Listeria innocua AL592022
ACCESSION AL596163.1 GI:16412421
VERSION AL596163.1
KEYWORDS Listeria innocua.
SOURCE Listeria innocua.
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Listeria.
REFERENCE
1 (sites)
AUTHORS
Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A.,
Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T.,
Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Dehoux, P.,
Dumann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L.,
Dussurget, O., Entian, K. D., Fsihi, H., Portillo, F. G., Garrido, P.,
Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J.,
Jackson, D., Jones, L. M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F.,
Kurapkut, G., Madueno, E., Maitournam, A., Vicente, J. M., Ng, E.,
Nedjari, H., Nordstiek, G., Novella, S., de Pablos, B., Perez-Diaz, J. C.,
Purcell, R., Rammel, B., Rose, M., Schlueter, T., Simoes, N.,

```

PARFPNLLVNGSSGIAGVGMATNIPTHHGLVEIDVGLALSHDPEITIRDMEXYIPGPDF  
PTAGIMRGSGIRAYSGSITVRGRVDIEKKNGKETIVITEIPQVANKARLVER  
IAELARKIGDITSLNDESDSGMRIVIEVRDLSASVIVNLFKMTATOTTEGINM  
LALYDNPKVLNLEKILLYLXLEHKVVRNTEFELPKAEARAHLEGLRALDNIDA  
IILKLRKSDVAKEGMLTFNLSKQALDMRQLRGLTGLEREKEEYQNLVAL  
INDUKAILADDERLEIRELEEKVKYADKRRETELAGDLVSTHEDDLTPEEVAI  
TLTKRGYIKRLPLSTYRSQRRGGIOGMSHEDDFVEHLVATSTHTDLTFLFTTKGV  
YRSKGYEVPYGRKAKIPIINLIGIESQEQVNAVILSEFTDDSYLFTTKHGVVKR  
TTLQFARIGRGLRAVELRENDELISVQMTDGSKNMIIATKHGSIVFPRENRVMG  
RTAGVRCIRLREDEVIGMEVLEDEKVLVYVTEKGYKQTPASQYPLNRGGMGVKT  
VTTEKGNLVAMKTVTGEEDMLMTVSGVLIRFEIDTVSQTGSRMSGVKULIRLDEDE  
KVATVAVKPEDEVELEEETIDETLITQVPDESFDAPGSDIE"

terminator

10600. 10653

/gene="gyrA"

10712. 12242

/gene="lmo0008"

10712. 10717

/gene="lmo0008"

10728. 12242

/gene="lmo0008"

/note="similar to cardiolipin synthase"

/codon\_start=1

/transl\_table=11

/protein\_id="CAC98223.1"

/db\_xref="GI:16409367"

/translation="MRKLIQFLAVLVFLVEYVLINQAAILFLVTSGIOLCGVIA  
IRLLFPORNTSSKVAVVAVIFILPVLTISYLVFGRNPATRKSTAQVKKTKLINA  
IHAIPNNTNEKPLRSKRAHLSITIEPIKNGKIELNGEETFPVLLDALRRKRNHII  
IOYIFTDALSTIRDLVEKAGSVEVRFMDGLSSKLGKAPLAPLKAGVSIHA  
FDPIASPIWATNLRNHRKIIVIDQGTGFTGLNIGIEEYRNTPDFRWRDTHIKIT  
QOAVTELQESPLNDWYVMEAGAGADGFISSGSKQVSPVDMGDWAQVYVGGPYDK  
EKWRDMLDLSAKESVTVSPYFVPDESLAVIRRVANSQVDRVILPKGDGRI  
SFGSNAYVTKMIAGAKMYAYADDSFVHAKMLVDGTRAAIGTANFDVRSFRLNHEL  
NVFLYDESEAHHLKDFKDFEDSRFLTKMDKNKPLLTIRIKEVLSLLSPIL"

## Alignment Scores:

Pred. No.:	5,48e+03	Length:	200050
Score:	51.00	Matches:	10
Percent Similarity:	65.38%	Conservative:	7
Best Local Similarity:	38.46%	Mismatches:	9
Query Match:	39.53%	Indels:	0
DB:	1	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AL591973 (1-200050)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20

: : : : : ||| ||| : : : : : ||| ||| ||| ||| : : : : : |||

Db 51231 GCTGAACAAAGAACAACTAAGTGATGAACAGATCGACTTGACCGAGAGATGACAGT 51290

Qy 21 ValValLysAlaAspAsn 26

: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

Db 51291 GGAGTTGAAAGAAAGAAC 51308

## RESULT 39

AL646065

LOCUS

DEFINITION Ralstonia solanacearum GM11000 chromosome, complete sequence;

segment 9/19.

ACCESSION AL646065

VERSION AL646065.1

KEYWORDS GI:17428522

SOURCE Ralstonia solanacearum.

ORGANISM Ralstonia solanacearum

Bacteria; Proteobacteria; beta subdivision; Ralstonia group;

Ralstonia.

REFERENCE 1 (bases 1 to 202050)

AUTHORS

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,  
Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L.,  
Chandler, M., Choisme, N., Claudel-Renard, C., Cunnac, S., Demange, N.,  
Gaspin, C., Lavielle, M., Moisan, A., Robert, C., Saurin, W., Schiex, T.,  
Siquier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M.,  
Weissenbach, J. and Boucher, C.A.

TITLE Genome sequence of the plant pathogen Ralstonia solanacearum

## JOURNAL

REFERENCE

2 (bases 1 to 202050)

AUTHORS

Boucher, C.A.

TITLE

Direct Submission

JOURNAL

Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CRPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS  
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,  
F31326 Castanet-Tolosan Cedex. Laboratoire de Genétique Cellulaire  
INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.

## COMMENT

FEATURES

source

1. 202050

Location/Qualifiers

/organism="Ralstonia solanacearum"

/strain="GM11000"

/db\_xref="taxon:305"

complement(210..686)

/gene="Rsc1507"

/note="RS03796"

complement(210..686)

/gene="Rsc1507"

/function="miscellaneous: unknown"

/note="Product confidence: hypothetical

Gene name confidence: hypothetical

predicted by Codon\_usage

predicted by FrameD"

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="HYPOTHETICAL PROTEIN"

/protein\_id="CAD15209.1"

/db\_xref="GI:17428523"

/translation="MEPRIDIPLEIFRTOLOALHTLTLOESQERMLKLTLSVEDA  
EORSLEACASAEAKTPWEMALPTSLQCRTEQNTKMFQWGLMNSQHALQOEAH  
TCADAWRTIASNMGSDPLMAPVRSFLDPLSETRPPPTAPAKKSSATATASAT"

complement(863..1255)

gene

/gene="Rsc1508"

/note="BS06024"

complement(863..1255)

CDS

/gene="Rsc1508"

/function="miscellaneous: unknown"

/note="Product confidence: hypothetical

Gene name confidence: hypothetical

predicted by FrameD"

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="HYPOTHETICAL PROTEIN"

/protein\_id="CAD15210.1"

/db\_xref="GI:17428524"

/translation="MPRAKLOVPAPPDNDNRNHRPAPPALASTTETTHAASVATR  
DAFSRRQLASVTPSTGTGTARGELGTGRAPSRMRSSANLAHAPARVAPGWPAATQ  
PLDGVRDSILAHMECALTYTISEAARS"

complement(863..1255)

gene

/gene="Rsc1509"

/note="BS04827"

complement(863..1255)

CDS

/gene="Rsc1509"

/function="miscellaneous: hypothetical/global homology"

/note="Product confidence: hypothetical

Gene name confidence: hypothetical

predicted by Codon\_usage

predicted by FrameD"

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="CONSERVED HYPOTHETICAL PROTEIN"

/protein\_id="CAD15211.1"

Tierrez.A., Vazquez-Boland,J.A., Voss.H., Wehland,J. and Cossart,P.  
Comparative genomics of *Listeria species*  
Science 294 (5543), 849-852 (2001)  
21537279  
2 (bases 1 to 200050)  
Glaser,P., Frangeul,L. and Rusniok,C.  
Direct Submission  
Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des  
Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris  
Cedex 15, FRANCE  
E-mail: pglaser@pasteur.fr  
Phone: +33 1 45 68 89 96; Fax: +33 (0)1 45 68 87 86.

## COMMENT

## FEATURES

source

Location/Qualifiers

1..200050

/organism="Listeria monocytogenes"

/strain="EGD-e"

/db\_xref="taxon:1639"

305..310

/gene="dnaA"

305..1673

/gene="dnaA"

318..1673

/gene="dnaA"

/codon\_start=1

/transl\_table=1

/product="Chromosomal replication initiation protein DnaA"

/protein\_id="CAC98216.1"

/db\_xref="GI:16409360"

/translation="MQSIEDIQETLQIVKKNKSPSYDTWMTKSTAHSELEGTFTIIS

ANNFVRDMLKSYQFTIANIQETGLRFDVDFDQGFENFETVIVKPNALDEBG

IEGRHVOQHNDKAVMYLSSEKFTNEFISSRDNRKTEFRTKYRNVDLLDDIOFL

AQKQTEFFETNTLYDEQKIIISDRPKPEITLEDRLSRFEWGLIIDITPPD

LETRAILKAKACGLDIPNEVMYIANGIDSNIRELEGALIRVVAYSLSVKNQDITA

GLAAALKDIIPSSKQVITISGIQFVGEYFHVLEDFKAKRTKSIAFPQJIAWYL

SRLETDASLPKDEEGGRDRTVTHAHEKISQLLTKDQVLNDLAEIEKNLRKAQNM

P\*  
1856..3062

/gene="dnaA"

1856..1860

/gene="dnaA"

1867..3012

/gene="dnaA"

/codon\_start=1

/transl\_table=1

/product="DNA polymerase III, beta chain"

/protein\_id="CAC98217.1"

/db\_xref="GI:16409361"

/translation="MAFVIERDLVQAVNEVTRATSTPIILTGIKIVNDEGVTL

TQSDSDISAEIPLIENDEIVEVEFSGIVLQSKYFGDIVRRLEPNEVEIEVTSNY

QTNISGQASFTINGLDPMYKLPVETDGTIKIPINLVKIVRQTVFAVAIEVRP

VLTGVNMIKENKLSAVATDSHRLALREIPLDIEDENYIVIPGKLSLAKLLDDA

SESIEMTANNOILFKDLDFYSLRLEGSPDTSRLIPTDKSELVINSKAFLOAID

RALLARENNVVIKMTLENGQVEVSSNPSGVNVSQSFGEIKISFNCKY

MDLALRFGDDIQLISFSGTMRPFLVRPKDAANPNLLIQLIPVRY\*  
3030..3062

/gene="dnaA"

3104..3109

/gene="lmc0003"

3104..4464

/gene="lmc0003"

3121..4464

/gene="lmc0003"

/notes=conserved

/codon\_start=1

/transl\_table=1

/product="hypothetical protein"

/protein\_id="CAC98218.1"

/db\_xref="GI:16409362"

/translation="MMKDMTGNPTKLIFLAMPMLIGNLFOQFTYMDAIVIGKFSV

VDAAAGVATNSVNFEMFISLIIIGMSGISVVAQYFGKDYDRLKDVATATVAVVFS

AIIITLVAGVLLAKPILLLRTPANILDSTIFLTLFGLPLMSLVNGMAAILRALGN

STIPLFLILSLMNLALDFLVYMDGVGRGAIAATVLSQFAAAIAVYIARHYVPF

MIRERAKFLSTPLAKEMVRIGLPSGLQGSFISIGNNALQSLINGFGSSVVAATTAAS

RBS

gene

CDS

RIDSITYQRIAGFAASSMFAQONIGAGKIDRVREGFWSGIKVWTAISIGITILVOLFF  
ARQFILLFYDSSETEVINIGSVYLLVSLFYVVGILFYVRETIRGTGDMVPLANGI  
FELVSRIVIGFISLYIGVGLWATPVAWITATILGVNRYKSGAQKKAVIRRK\*  
4631..4637  
/gene="lmc0004"  
Join(4631..4637,4644..4865)  
/gene="lmc0004"  
4644..4865  
/gene="lmc0004"  
/note="similar to B. subtilis YaaA protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC98219.1"  
/db\_xref="GI:16409363"  
/translation="MAETVKINSEFVTLGQLQLLQIMDVSTGGMAKAYLSENTIYINGE  
QDNRRGKLRNGDVLIVPGVGKVIQGRK\*  
4856..4861  
/gene="RecF"  
4856..4861  
/gene="RecF"  
4869..5981  
/gene="RecF"  
4869..5981  
/gene="RecF"  
/codon\_start=1  
/transl\_table=1  
/product="RecF protein"  
/protein\_id="CAC98220.1"  
/db\_xref="GI:16409364"  
/translation="MHLESIVLRNFRNYENLELEFSPSYNVFLGNAQGNLLEAVL  
MLAKASHNTNDKDFIMWEKEAKMEGRKAGHQSVPLELAIQKGRKAKVNHLEOK  
KLSQVYGNVNVIFAPEDISLVKAGFIRRELNMEIGMQPIYHLNSEYOHLEOK  
NOYLMKQNRKVDPTLLDITLQEFADVAINTLKRRADEIQLEAYAAATHHQISRL  
ETLKIYKASITLNGDDPEWKRADLQKMEISKQRIDRGVTLIGPHRDSLSFYINGQ  
NQDQSGSQGQRTTALSIKIARITLHEETGEVPLLLDDVLSELDDYRQSHLGAIE  
GKVOIFVTSTSGIDHETLKQATTFYVEKGTVKKS\*  
6017..7970  
/gene="gyrB"  
6017..6021  
/gene="gyrB"  
6030..7970  
/gene="gyrB"  
/codon\_start=1  
/transl\_table=1  
/product="DNA gyrase subunit B"  
/protein\_id="CAC98221.1"  
/db\_xref="GI:16409365"  
/translation="MSEENITNVQENASDYNEQIQVLEGLAVRKPGRMYIGSTSOR  
GLHVLVWEIVDNAIDALAGFCTEIEITIEDNSITVRDNGRGIPTGINKIGRPTVE  
VIFTVLHAGKFGGKGYSGGLHGVGASVYNALSTSLVYVHREGQKYVORFERGDV  
VMDMEQGDYRGITIVHETPDQIFTEFEDFTLRTRELALNRGLTISIEDK  
REHVKRKDFHVEGGIRSYVHLNKARDVIHEPIYLEGEDDIMVEISMQYNTGFS  
NIISFANNTHVEGGTHESGFKTALRVINDYARNKLKDSDDNLSGEDVREGTLAI  
ISLKHDPQEGQTKTLGNSEARSITDKLSEALNFMENPDVAKKIVKEGVASR  
ALAKARAREVARKSSGLSESSLPGKLDCSSRNPELSLYIVEGDSAGSKAQGRDR  
LQAALIRKILNVEKARDIRILANEINTIFAMTGFGGFDVSKSYHKLIMT  
DADVGAIHTLLTLFIRYMRPLDAGIYIQAQPLYQIKHGKQLEYVSDGLEDY  
LASLDGDTKYSIQRYKGLGEMNPEQLMDTMTNPEHRTLLQVNIKDAIDADETFEMLMG  
DRVEPRRKIEDNAQVVKNDV\*  
8048..10653  
/gene="gyrA"  
8048..8053  
/gene="gyrA"  
8065..10593  
/gene="gyrA"  
/codon\_start=1  
/transl\_table=1  
/product="DNA gyrase subunit A"  
/protein\_id="CAC98222.1"  
/db\_xref="GI:16409366"  
/translation="MAETPNORITEINLNKEMRTSLDVAMSIVIVARLPDVRGLKLP  
VHRRLIYAMNDLGMTSDKAYKKARSARIVGEVYHPHGDATVFTVWRNAQDYSYRNM  
LVDHGNFGSVGDGMAAMRYTEARMSKISMELLROINKDITIDYADNIGSREPEVIL

RBS

gene

gene

CDS

gene

RBS

CDS

gene

RBS

CDS

terminator

RBS

gene

CDS

and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

#### FEATURES

source  
1..175115  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
/map="22C-22C"  
/clone="BACR1F10 (D1463)"  
/clone\_lib="RPCI-98 (Roswell Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
pBACe3.6)"

BASE COUNT 49696 a 39203 c 37955 g 49261 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4.77e+03 Length: 175115  
Score: 51.00 Matches: 11  
Percent Similarity: 53.85% Conservative: 3  
Best Local Similarity: 42.31% Mismatches: 12  
Query Match: 39.53% Indels: 0  
DB: 3 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC105899 (1-175115)

Qy 2 AspaLaLeuGluaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21

Db 148 GATCGCGGGAATCATATAGACACACAGCAGCAGAGGAGGAGGAGGAGGAGGCG 207

Qy 22 VallysAlaAspAsnAla 27

Db 208 GTACGAGGAGATCGTGCC 225

#### RESULT 37

AC092244

#### LOCUS

DEFINITION Drosophila melanogaster, chromosome 2L, region 22B-22C, BAC clone

AC092244

AC092244

HTG.

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

AC092244

McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Direct Submission  
Submitted (30-JUN-2001) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US

Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

#### FEATURES

source

1..183898

/organism="Drosophila melanogaster"

/strain="y; cn bw sp"

/db\_xref="taxon:7227"

/chromosome="2L"

/map="22B-22C"

/clone="BACR19N22 (D1355)"

/clone\_lib="RPCI-98 (Roswell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

pBACe3.6)"

BASE COUNT 50623 a 41508 c 41071 g 50696 t

ORIGIN

Alignment Scores:

Pred. No.: 5.02e+03 Length: 183898

Score: 51.00 Matches: 11

Percent Similarity: 53.85% Conservative: 3

Best Local Similarity: 42.31% Mismatches: 12

Query Match: 39.53% Indels: 0

DB: 3 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC092244 (1-183898)

Qy 2 AspaLaLeuGluaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21

Db 130591 GATCGCGGGAATCATATAGACACACAGCAGCAGAGGAGGAGGAGGAGGCG 130650

Qy 22 VallysAlaAspAsnAla 27

Db 130651 GTACGAGGAGATCGTGCC 130668

RESULT 38

AL591973

LOCUS

DEFINITION

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Direct Submission  
Submitted (30-JUN-2001) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US

Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

FEATURES

source

1..183898

/organism="Drosophila melanogaster"

/strain="y; cn bw sp"

/db\_xref="taxon:7227"

/chromosome="2L"

/map="22B-22C"

/clone="BACR19N22 (D1355)"

/clone\_lib="RPCI-98 (Roswell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

pBACe3.6)"

BASE COUNT 50623 a 41508 c 41071 g 50696 t

ORIGIN

Alignment Scores:

Pred. No.: 5.02e+03 Length: 183898

Score: 51.00 Matches: 11

Percent Similarity: 53.85% Conservative: 3

Best Local Similarity: 42.31% Mismatches: 12

Query Match: 39.53% Indels: 0

DB: 3 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC092244 (1-183898)

Qy 2 AspaLaLeuGluaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21

Db 130591 GATCGCGGGAATCATATAGACACACAGCAGCAGAGGAGGAGGAGGAGGCG 130650

Qy 22 VallysAlaAspAsnAla 27

Db 130651 GTACGAGGAGATCGTGCC 130668

RESULT 38

AL591973

LOCUS

DEFINITION

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

AL591973

```

/protein_id="AAK15440.1"
/db_xref="GI:13194222"
/translaton="MAGASAKVAMLLSVLATYGFAGVYVYVNDLWLPKATWYGQNG
AGPDDNGGCFKNTNYYPPMSMTSCNEPLFDQKCGCACYQIRCTNNPSCSQPHT
VIITDMNYGVARYHFDLSGAFAMARPGINDQLRHAGITDQFRVPCYHRLGYVN
FHYEAGSNPYLAVLEFAKNDGVVQLDVMESLPSCKPRTVMTPMRRSWSIWLRLDA
NHRLOGPFSLRWVSESQTVIAQVIPANWRANTYNGSKVQPR"
repeat_region
complement(1312..17335)
/rpt_family="(CA)n"
/rpt(21858..22143,22248..22357,22530..23373)
mRNA
/gene="OSJNB0014111.3"
21858..23373
/gene="OSJNB0014111.3"
/note="identical to beta expansin EXPB6 GB:AAF72987
GI:8118430 (Oryza sativa) (Proc Natl Acad Sci U S A
94 (12), 6559 6564 (1997)); EST AU062937, C72255, AU161321
from this gene"
CDS
join(21954..22143,22248..22357,22530..23057)
/gene="OSJNB0014111.3"
/codon_start=1
/product="beta-expansin EXPB6"
/protein_id="AAK15442.1"
/db_xref="GI:13194224"
/translaton="MAAMGSKVAAILLILSVLVHSGKGHVYVYVNSDASAYGSH
LPARATWYAGTPAGPDNGAGCGFRKNYQIPFSSMTSCGNEPIFKDGKCGSCYQIR
CNKDPSCSGNIETVIITDMNYVPARYHFDLSGAFAMARPGINDQLRHSGITDIQF
RRVPCNPGKINIFHVEGNSPNVFAVLVEYEDLDGVDVQVQVDMLESKSYGGATGVMT
PMRESGSIWRLSDNHRLOAPFSLRISDSGKTLVANNVIPANWSPNSRYSIVQFS"
23102..23121
/rpt_family="(CGAT)n"
complement(join(26118..26245,26118..26381))
/gene="OSJNB0014111.4"
/note="predicted by fgenesh"
complement(join(26118..26245,26351..26381))
/gene="OSJNB0014111.4"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK15444.1"
/db_xref="GI:13194226"
/translaton="MPEKQAPKPAVAAPQRGGRGCAADGVHERAGDRPKAQWM
DGWRREG"
complement(26200..26222)
/rpt_family="GC-rich"
complement(27602..27629)
/rpt_family="GC-rich"
27980..28118
/rpt_family="putative stowaway type MITE element"
complement(27991..28129)
/rpt_family="putative stowaway type MITE element"
complement(join(29835..30595,31066..31120))
/gene="OSJNB0014111.15"
complement(29835..31120)
/gene="OSJNB0014111.15"
/note="predicted by gensecan"
complement(join(29835..30595,31066..31120))
/gene="OSJNB0014111.15"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK15452.1"
/db_xref="GI:13194234"
/translaton="MKEKERKADSAVLGFPRAMERERTTPSAATSLHHRRLTLVD
AIGSKAVCGGERGERRIQPHRPSLFRHQACQRPVGAERFPFPFHSDDGTIM
RPFPVRIIAALRPGPLPRRAARLPHLNPVPPPPPPASKLPIILYFHGSGFLVFT
DNVYFHASCEANAAATVPAILVSLDYLRAPEHLPAAYVSTVWLCDAAAGDPWLAASA
REVAGVAVAGCGLALSSLPVSGRHHSSPPPLPLPAQAQEREREGDREKW"
complement(29843..29879)
/rpt_family="(GA)n"
30600..30652
/rpt_family="(CGACA)n"
complement(32120..32144)
/rpt_family="AT-rich"
join(<33752..33943,35026..>35142)
/gene="OSJNB0014111.13"

```

```

Alignment Scores:
Pred. No.: 3.66e+03 Length: 135509
Score: 51.00 Matches: 13
Percent Similarity: 57.69% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 11
Query Match: 39.53% Indels: 0
DB: 8 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AC037426 (1-135509)

Qy 3 AlalaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 22
Db 67185 GCGCGGAGGCGGCGGCTACGGAGGACGACGCGCGCTACGGAGGCGGCGACGACGCG 67126
Qy 23 LysAlaAspAsnAlaAa 28
Db 67125 GCGCGGATACGGCGCG 67108

RESULT 36
AC105899 175115 bp DNA linear INV 10-JAN-2002
LOCUS Drosophila melanogaster, chromosome 2L, region 22C-22C, BAC clone
DEFINITION BACR11F10, complete sequence.
ACCESSION AC105899
VERSION AC105899.1 GI:18104818
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 175115)
Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Fierler, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Scheeler, F.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 2L, region 22C-22C
Unpublished (1998)
2 (bases 1 to 175115)
Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Fierler, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Scheeler, F.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Direct Submission
Submitted (10-JAN-2002) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location

```



(657 aa), fasta scores; opt: 479 z-score: 421.7 E():  
3e-16, 28.2% identity in 333 aa overlap. Contains possible  
N-terminal signal sequence"

/codon\_start=1  
/transl\_table=11

/product="putative secreted protein"

/protein\_id="CA20497.1"

/db\_xref="GI:3451438"

/db\_xref="SPTRMBL:O86543"

/translation="MEANVPTVLTPLKARLLLVALLVLTPTAPEALAAGKPETGA  
TRHAEVAVTRADVDLTSTALGGRVTEVRLTDPGNPHDRHQRHPTLWL  
LHCCGDTYTSWTQDVAETSLRDVLVNPAGWNGWYSDWNHGGGDPANETPTHT  
KEHLHLERDNGAGSRVAGLSKGGGALLIYAARHPGMPRTAARSGSAHPLLNDES  
VDRIHGAGDNDPLRWGDPVQAQRIGWADHPFLAKRLRSIPVYLSGDGGTTGPL  
DAPGATGDDNDRNQNHAAELKRVGARHVTTHFYGPCTHGWAYRERELHASLPML  
LGALRVGD"

6183. 6186

RBS

Alignment Scores:  
Pred. No.: 784 Length: 30561  
Score: 51.00 Matches: 12  
Percent Similarity: 60.00% Conservative: 3  
Best Local Similarity: 48.00% Mismatches: 10  
Query Match: 39.53% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x SCLF2 (1-30561)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAa 20

Db 15463 GCGGACGGCTGCGAGCCCTCGGACAGATGCGGGCCACCCATTCGTCGGTCCGTCGAC 15404

Qy 21 ValVallysAlaAsp 25

Db 15403 GTGATCGCGCGGAC 15389

RESULT 33

AC005749/c

LOCUS

DEFINITION

Drosophila melanogaster DNA sequence (Pls DS06378 (D278) and

DS03995 (D277)), complete sequence.

AC005749 AC004576 AC004575

AC005749.1 GI:3688071

HTG.

Drosophila melanogaster (Subclones in sac from P1 clones DS06378

(D278) and DS03995 (D277)) DNA.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 96288)

Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,

Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,

Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,

Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,

Kim,S.H., Lee,B., Lomotan,M.A., Mak,J.J., Mazda,P., Moshrefi,A.R.,

Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., PUNCH,E.,

Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.

Sequencing of Drosophila chromosome 2L, region 22B3-22C1

Unpublished (1997)

2 (bases 1 to 96288)

Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,

Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,

Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,

Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,

Kim,S.H., Lee,B., Lomotan,M.A., Mak,J.J., Mazda,P., Moshrefi,A.R.,

Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., PUNCH,E.,

Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.

Direct Submission

Submitted (02-OCT-1998) Berkeley Drosophila Genome Project, MS

64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,

Berkeley, CA 94720, US

Sequence submitted by:

FEATURES

source

1. 96288

Location/Qualifiers

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/chromosome="2L"

/map="22B3-22C1"

/clones="Pls DS06378 (D278) and DS03995 (D277)"

/note="These two Pls were sequenced as a project. DS06378

(D278) extends from p1 end at bp 1 to p1 end at bp 85081.

DS03995 (D277) is a bridge clone extending from p1 end at

bp 57,688 to bp 96,288 minimally overlapping its proximal

neighbor DS00164 (D269)."

BASE COUNT 26399 a 21521 c 21934 g 26434 t

ORIGIN

Alignment Scores:

Pred. No.: 2.57e+03 Length: 96288

Score: 51.00 Matches: 11

Percent Similarity: 53.85% Conservative: 3

Best Local Similarity: 42.31% Mismatches: 12

Query Match: 39.53% Indels: 0

DB: 3 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC005749 (1-96288)

Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAaVal 21

Db 49770 GATCGCGGATCATATAGAGACACACAGACGAGAGAGACGAGACACACGCGGCG 49711

Qy 22 VallysAlaAspAsnAla 27

Db 49710 GTACGAGGATCGTGCC 49693

RESULT 34

AC019871

LOCUS

DEFINITION

Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*

pieces.

AC019871

VERSION

AC019871.1 GI:6665026

KEYWORDS

HTG: HTGS\_PHASE2.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 105823)

Adams,M. and Venter,J.C.

Submitted (30-DEC-1999)

Direct Submission

Rockville, MD, USA

This sequence was identified as CDM10211226 by the submitter.

For more information on this record e-mail to fly@celera.com.

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

FEATURES

source

1. 105823

Location/Qualifiers

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT 29823 a 23528 c 23181 g 29291 t

ORIGIN

Alignment Scores:

Pred. No.: 2.57e+03 Length: 96288

Score: 51.00 Matches: 11

Percent Similarity: 53.85% Conservative: 3

Best Local Similarity: 42.31% Mismatches: 12

Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://fruitfly.berkeley.edu/sequence/>) or send  
email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

Library locations: 42-67, 59-42.

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, ggg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 1F2 lies between 8A6 and 4G2 on the AseI-A genomic restriction fragment.

## FEATURES

## source

1..30561  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"

1..936  
/clone="cosmid 1F2"

1..102  
/gene="SC1F2.01"

## misc\_feature

1..927  
/gene="SC1F2.01"

/note="nominal overlap with cosmid 8A6 from 37344 to 37445"

## misc\_feature

1..927  
/gene="SC1F2.01"

/note="Pfam match to entry PF00232 glycosyl\_hydrol, Glycosyl hydrolases family 1, score 350.20, E-value <1..936"

## CDS

1..927  
/gene="SC1F2.01"

/note="SC1F2.01, probable beta-glucosidase, partial CDS, len >311 aa; similar to many e.g. BGLS\_AGRSP beta-glucosidase (EC 3.2.1.21) (459 aa), fasta scores: opt: 784 z-score: 986.7 E(): 0, 42.5% identity in 315 aa overlap. Overlaps and extends SC8A6.25. Contains Pfam match to entry PF00232 glycosyl\_hydrol, Glycosyl hydrolases family 1, score 350.20, E-value 2.2e-101"

/codon\_start=1  
/transl\_table=11

/product="putative beta-glucosidase"

/protein\_id="CAA20492.1"

/db\_xref="GI:3451434"

/db\_xref="SPTREMBL:O86539"

/translation="AEHFAAYVAERLGDRAVHATLNEPLCSNIGHLEKKNAPG QTDLRNRAVSHELLHGLMAQALRAAPDAQVIGVNNLSITHAATDREEDLAARR MDGHTNRWLDDVHGFQPDQMRVEYGVLPDPGRMLGAELDWLGLNYFPQTV AADTPAPQVPRVRAGVPQDMDEIDAGGTEELLRLTEYGPRLKLYTENGSSFP DLVRPDGVDDPERQEYLTALAACASARKGAPLAGYFAWSLLDNFENWAYDYKRFGLVHVDYATQVTRTKSGHHYADIIGRAGRERKAA"

1051..1236

/note="SC1F2.02, probable beta-glucanase pseudogene fragment, len: 61 aa; this appears to be a gene remnant similar to part of e.g. TR:O68641 (EMBL:AF052745) beta-1,3-glucanase II from Oerskovia xanthineolytica (435 aa), fasta scores: opt: 100 z-score: 207.7 E(): 0.00025, 38.3% identity in 60 aa overlap. There is no start codon"

/codon\_start=1

/pseudo

/transl\_table=11

/db\_xref="PID:e1316894"

complement(1275..2177)

## CDS

/gene="SC1F2.03c"  
complement(1275..2177)  
/gene="SC1F2.03c"  
/note="SC1F2.03c, unknown, len: 300 aa; similar to upstream gene SC1F3.04, E(): 0, 48.8% identity in 260 aa overlap"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein SC1F2.03c"

/protein\_id="CAA20494.1"

/db\_xref="GI:3451435"

/db\_xref="SPTREMBL:O86540"

/translation="MMFRPALTCQVSTLEPWAHELRTESGRMTSLSSGFAAHPDRI DTSRPHAPMYDYFLGGKNDYEDQRAAQPKSAPEVREGVLANRHFMRHVRHVVA EGVROILDVGTGLPTPEPNVHQIARSVAFTGVVNDQPIVATHSRALADPDPTAVV LADLRDRPRILEHPEVRAVDFDRPVALLVAVHVFVADAQDPAGTAVTLRALDPAGS YLVLSHATGDVHEDRREDAAYNVKATASLNLRSHAAVLDFGDFSLVPEGLVRVTDW RPEEARPPDAPPIGIYGVARKDA"

2299..2304

/note="possible RBS upstream of SC1F3.04"

2308..3093

/gene="SC1F3.04"

2308..3093

/gene="SC1F3.04"

/note="SC1F3.04, unknown, len: 261 aa; similar to upstream gene SC1F2.03c, E(): 0, 48.8% identity in 260 aa overlap"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein SC1F3.04"

/protein\_id="CAA20495.1"

/db\_xref="GI:3451436"

/db\_xref="SPTREMBL:O86541"

/translation="MTAGTPRPDIDTSKPHPARVYDMLGGKNDYDQVGVGTLPAAE AGNARNRPAFMHRAAAWLAQNGVDFLDIGTGIPTEPNLHQIQAISPAARIVYADN DTVLRHAEALLTGPPEGVYDFLHADVRQPATILERAREVLDFDRPIALSLLHFL PDDEPYRVVSTLVDTLPRGSYLLSHGTADQHPKQFTESAYKKGATPLRMRTSE VSPFCGLELVKPLGVFATVQEEPAVPRESGFYAGVARIP"

3223..4881

/gene="SC1F2.05"

3223..4881

/gene="SC1F2.05"

/note="SC1F2.05, probable secreted protein, len: 552 aa; contains N-terminal signal sequence"

/codon\_start=1

/transl\_table=11

/product="putative secreted protein"

/protein\_id="CAA20496.1"

/db\_xref="GI:3451437"

/db\_xref="SPTREMBL:O86542"

/translation="MKLRKRRLCAATGVVLGALTALALPPAAVASPHPGGHPPTARR ALPLGPAGLPERRTTTTLQGVTLTRVNGAEDPALHWTVEVSIPIGGTISPDQAPP TALKDRASADELAELRRDGFARVEEVTTPATADYRGTLGMVRVIGTDSQTAATA BERTLRAGYSGSAVYVWDGAATDRGPMHVDVLTIDPRTFRGLTDSAGYCPLENRET TSALAAASAAATVAGNAGFVLDPRAGAGDPAGCVGYDCRLLSEPAGRPALVVDHG RTEARTLTWQRIATRAASPLNGVNRVPLIRNCGGSAGDPTPTSLPHDVTCTNPD ELVATPDYGGRTPGEGMEAVLDAHERVVELRSRPGGTIPRGSSVQNTGERVADLT ATARVGERVYATLDDDRMRMSTPGDITVAGNELVRDGRVHVHTPATDGMVHPGD PSWYGTWIKRNPRTVAGDAAGTAVLTADGRSTDLSGLSIGESAEVARSGLRDAY NLDGGSTVHVAEAGVANSFSDAAGERPVGDALLILPHRHRTS"

4908..4991

/note="12 degenerate repeats of ACGCAC(C/A)"

5005..5008

/note="possible RBS upstream of SC1F2.06"

5014..6042

/gene="SC1F2.06"

5014..6042

/gene="SC1F2.06"

/note="SC1F2.06, probable secreted protein, len: 342 aa; some similarity to secreted antigenic proteins 85A, 85B and 85C from Mycobacteria e.g. A85B\_MYCLE (327 aa), fasta scores: opt: 325 z-score: 247.2 E(): 1.6e-06, 28.5% identity in 319 aa overlap and to secreted antigenic protein PSI from *Cornibacterium glutamicum* CSp1\_CORGL

## CDS

/note="SC1F2.02, probable beta-glucanase pseudogene fragment, len: 61 aa; this appears to be a gene remnant similar to part of e.g. TR:O68641 (EMBL:AF052745) beta-1,3-glucanase II from Oerskovia xanthineolytica (435 aa), fasta scores: opt: 100 z-score: 207.7 E(): 0.00025, 38.3% identity in 60 aa overlap. There is no start codon"

/codon\_start=1

/pseudo

/transl\_table=11

/db\_xref="PID:e1316894"

complement(1275..2177)









```

/transl_table=11
/product="Vng6208c"
/protein_id="AAG20870.1"
/db_xref="GI:10584263"
/translation="MKSQQLKGVAVGAFVLYLVVGGGLAWNPVQEGNIEVVKEMGA
STGTLEFGANVIVPIKQSTAVVPRPOEYTMANEKQGEARDDSVLEVLNDGVSVN
VDVIRVYNKTEAATPVYKDYQSOAEARLIRPTQDVLRTGEGDIDTTELYTGAGQ
KOMAAYKKALETEAVGSLIEAVQIARNIKLPGYADAVKEKVEKONIEKKNSIQ
VAKAEARKVQAKGEARENEIVAESLKNPELIKIRYIEALNQDNSTIYVAGASQAG
ITLTKVDDEENKS"
1016..1159
/gene="VNG6209H"
CDS
1016..1159
/gene="VNG6209H"
/feature="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng6209h"
/protein_id="AAG20871.1"
/db_xref="GI:10584264"
/translation="MAGRLIVFGLLWLGAIALVTVGAPRYFSERQRLSHEERMYDKRH
DDE"
1543..2808
/gene="gabT"
CDS
1543..2808
/feature="gabT"
/codon_start=1
/transl_table=11
/product="gamma-aminobutyrate aminotransferase"
/protein_id="AAG20872.1"
/db_xref="GI:10584265"
/translation="MAYTHAEGIEVTDAGGTYIDCLAGACTLALGHNHNPVATBLER
LLAEDRLHTDITTEPEAEVDALFDSLSPASPANTAKVQFCSPAGTDAIAAALKLT
TATGNSRVLAGHGYHGTAGALSMDIDAKEPVSGLMPNVHPLPYDPQYRHPFGLP
AGEHSITYIEHAISPERGFTPEAAVLEVOGEGVNPAPDAWQELRRIITREHD
IPLVDEITQGLTGETWSEFHAGIEPDVVTCSKAVGGSLPLAVVLYNEGDLWEPG
AHCTFFGNQLMAGARVTIEHVIEQDLPAHADKTGARLSALTDLEADFVQGVNVRG
RGLMGEVFPADPAAGDERPPDGLAAAVKQACFERGLIIEETGGRSAVAVLPPLI
VSSHQIDIEIAAFSDAVTAVLSSSAEVS"
2919..4379
/gene="bdb"
CDS
2919..4379
/feature="bdb"
/codon_start=1
/transl_table=11
/product="L-2,4-diaminobutyrate decarboxylase"
/protein_id="AAG20873.1"
/db_xref="GI:10584266"
/translation="MEACADAVLDVIGSAHGPGYSGASYDRLRLADTTALPAEGHPL
EDVLAAREVDYLANAVHPSDETCVAHLQCPMPVPAALAAEALLTASNQSLDSFQAPAA
TVVERLIADLTALVELGPAAGDVVTGGTETSNHQAALLARDDVETVFGVSREHGL
PPAAQDLRLCSADAHETAQSAALLGLGEDAVYIPTDGRHRMDATALRETVERLND
AKRIPFVVGATGTDGSDIPLSAVDAAVATHDLVHVHDAFGALAVSQHSERLS
GIDRADSVAVDFHKLFTPTQIACGALLVADGESFELMSNAAAYLNPAGDAVPLNAKST
RTTRDALPKYVAFRTFGRDGLGALVDTLALADDVAGLLRADPAFELACEPTLNAY
TFRRPVREHAHCDPGPWADHVTEAARQLFDAGTAVVARTTVDDRAHVKTTLMNPRT
TVADIQSVLVAFKGAHAATIEAADRTAPITTAAGGQ"
4376..6199
/gene="Iuca"
CDS
4376..6199
/feature="Iuca"
/codon_start=1
/transl_table=11
/product="iron transport protein A"
/protein_id="AAG20874.1"
/db_xref="GI:10584267"
/translation="NTGVGAVADTPAAHAESATLHAFLNCLYLRETDAGEVVSRAVPD

```

```

GVPAGACADRVRVVRPFPAASTTVFAPLRVASATGRHLFVAVPAPVAVPCVPDPACTP
SPIDAAGVASLARRELA.LDGDPTAGROLLSERVLASQTATPERFVAARDTPRDERLYD
TVTSFRDAEQSLVYGHMHPTNSRTGIPPEQSRYPAPELRAGFOLRVFADPALVET
WAAGDASATEQLSVVVVADATVPPESEATLVVHPWQADY.LHAQPHVAAA.TARDELRD
LGFAGPTPTPTSSVRLTWSPDAPFMYKGLAVETITNAERTNKRPELARGVAVSELAT
GFGDAFDRFPFAFSTIEDPAATLT.LGDCAESGFEAVLRNPFAAGDDLYDQVCPVLAALC
QGLDGPAAARATVSR.IATREGRTESAVAREWFOYLAHTVPPVWMLPVNOCVGLGAH
QONTLVR.LDDDCRPTAGFYRONQGYFFPASOAADVDMLPGVCEADTVCPDAIADER
LRYIVFNNAFGVYNALGAGGVYDERALLAVLRTLES.LAEHEPADSTLVLSALLTERR
VPCKANLLTRFEEREDELEADLETQSYTAIENPLVTRVDTQ"
6196..6906
/gene="lucB"
CDS
6196..6906
/feature="lucB"
/codon_start=1
/transl_table=11
/product="iron transport protein B"
/protein_id="AAG20875.1"
/db_xref="GI:10584268"
/translation="MTAPLPRITSYAYRRPQPDTRTVAFRRVALDRDLGRLLHAWLT
SDHVTYWSLGAFLPAFQAIGDHLDDHVTAYVGCIDGVPMSYVRYWVAADPLADH
YDADDNQG.IHILIGPPEYLGQCGCTALLRAMTALAFRHAATDRVVAEPDATNDAA.IA
AFERAGFPRETFREPHEDKDATRVVCDRERFVGAFSPCA.IAANGPPADGPAVNHDDA
THRCPSGSGNSTLEADR"
6903..8336
/gene="hxyA"
CDS
6903..8336
/feature="hxyA"
/codon_start=1
/transl_table=11
/product="monooxygenase"
/protein_id="AAG20876.1"
/db_xref="GI:10584269"
/translation="NTESVHDLVGLGPNLGAALLEGSDDVPLDAAFLBQKPEFY
WHDGMLIGTTLLEVFLADVTLADPTNPYSYLNLYLRENDRLRYEYFFTFQFLLPREY
NDGCMVASELAAPRFSRRVVDITAI.DDADTANFRVTAVEPADSEARSDDDADPTHQ
YRADLVYGVGSRPYVPPAFRDLPAQFPHFHTAAYREHRAAVCDADSVAVIGSGQAAE
VHDLTLRRSTSDRSRVDWTRPSAGFFPMYKSLGLQHFTEYDYVVDLPQVYKDE
LIPNODILYKGIDPDETSTRIDYTLVEQSGRDPNVGMIAATEVGAPEPTGGCRVVD
CRHQTDTGAFETDAVILGTGYHRPTPGFLSSLEPRLARODQDLAVDREYGVGSDL
EGGLFVQNELTHGVGAPDGLGAYRNAVILDRLLADAPYPVDRDTVFQDPAPAPAYA
SHAPGGRVDRDDPTTPPAAGQPNHHQ"
8362..10188
/gene="lucC"
CDS
8362..10188
/feature="lucC"
/codon_start=1
/transl_table=11
/product="iron transport protein C"
/protein_id="AAG20877.1"
/db_xref="GI:10584270"
/translation="NNTYEDTLQDAVTADTWTATVERSLLAKLLIREFTYEELIAPERTT
DTRGRYRLVLPAA.DDADTVRYRFEAEERLMDTRVFPDTIERAPADSWATATDP
TRFLLDARTELIDATETLTTGHLIREYRNTLLADAH.IEARSAPESFADCSVAYESEMVG
HPWLTFNKGRLGMDYEDYAPERRQTTTICMLAVRTDRVTFVSGITPTGILLRGE
LQDQVDFMATLADKGLDQPEYTLPLPVNYQWETTITSYASELADALVPLCNGADS
YLPQSVRTLFNVDAPERINVKLP.MKIRNTLVWGLPGDRTEAARVTEYITDIDRGG
PFLRDCDELVLPGVAGVNYDHPAFSELEDNAYQYDELLGAVRESDITDA.IFDSQAV
TLAALLHEADGTAVISERVSLSLAWLTFLDTLLPLLHYLYRYGVTFSPHGEN
TIVILDEEFVPTRLGKDFVDDVNI.SDQPLPELADLADLRVLASEPNGLCQFVFC
GLFVGVRYLSDVLTAAHGYDETEWTVQVRAITAYQAEFPELSDRFETFDLLQPTFT
KLC.LNRRLTEGYS.DTASRRPAAHCHGTVTNALHEVAGDADE"
complement(10189..10842)
/gene="trkA3"
complement(10189..10842)
/feature="trkA3"

```

```

CDS
79..1992
/gene="CG7289"
/note="Longest ORF"
/codon_start=1
/db_xref="FLYBASE:FBgn0031375"
/product="Ld12115p"
/protein_id="AAL28694.1"
/db_xref="GI:16768950"
/translation="MLFRSAKSGFDAGDVMRRKRLAEGTSSRYRLLFLLLCVCVVF
YMPITFRYIFLSAPEQKDPHSCMDRLTPFIONFEDANIRHVSAPKMPGERDFT
PVGNGYGLGTAHDAFLNKNRAWOLPFPQPVSVSGSASGKERATVVEYLTG
MYHRCQCFAGVFVSYAHRTQPNIFHQEOLITNTRNLLEDIELMPRVNLQKLTRR
TVLSEPVSGVFTTTELEVSQTVOLQAEKPSKIVISIVKQMDSKLRLKRGTVR
IYPTAVHVSQVAAEIKGIGSETIEQQATQAMAKLILQKSGPNPGLSNVNIYRQ
EHIDVTDLWAGFTISTSKAENSLNGDRINATMVAVLSQVRSFFEEAGGSKKEDIA
KALTYAEGCYDSYHTLOAENLWRENSLQOLNSLVTSMWLTLEKQCHNLIRAGASGV
IOAMVLSGCSFRESNOHLECNTHPKLFHEDFHERLNYGKTHVAVTIIIVDDDNKAVI
NIALDRSDSYACDGGCLDEPVLIIQNRQFPVKLTPELPAIILYITEDKQMHLELHH
ALHVEVSPAHEQHLIALHRHGHQGLPLTFWVSVCALIIIVHFIFLCKLIIEKY
EFSDKLRYRNKP"
BASE COUNT 586 a 558 c 516 g 491 t
ORIGIN
Alignment Scores:
Pred. No.: 50.4 Length: 2151
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
DB: 3 Gaps: 0
US-09-847-539a-6_COPY_59_86 (1-28) x AY061146 (1-2151)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
Db 228 GATCGCGGAATCATAGAACACACACAGCAGCAGGAGGAGGAGCAACAGCGCG 169
QY 22 ValysAlaAspAsnAla 27
Db 168 GTACGAGGATCGTGCC 151
RESULT 28
PFAS47
LOCUS
DEFINITION
Plasmodium falciparum (isolate NF7) S antigen gene, complete cds.
ACCESSION
M10130
VERSION
M10130.1 GI:160670
KEYWORDS
S antigen.
SOURCE
P.falciparum (isolate NF7) DNA, clone NF7.S.
ORGANISM
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
1 (bases 1 to 3838)
Cowan,A.F., Saint,R.B., Coppel,R.L., Brown,G.V., Anders,R.F. and
Kemp,D.J.
Plasmodium falciparum
Conserved sequences flank variable tandem repeats in two S-antigen
genes of Plasmodium falciparum
Cell 40, 775-783 (1985)
MEDLINE
85176931
COMMENT
Twenty degenerate tandem repeats of twentyfour base pairs each
start at position 953 and end at 1432.
FEATURES
source
1..3838
Location/Qualifiers
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
665..733
sig_peptide
665..733
/note="S antigen signal peptide"
CDS
665..1594
/note="S antigen precursor"
/codon_start=1
/protein_id="AAA29758.1"
/db_xref="GI:160671"
/translation="MNRILSVSYFLFLYLYIKTYGKVKNTDQELSNLYGNTNYLRN
GFLNGKNGKNTYDQEREGENDDEHSESDNDENIIVGQDEAPKSDAEAL
LKSDEAEALKSDEAEARKSDEAEALKSDEAEARKSDEAEALKSDEAEALKSDEAEARK
SDEAEALKSDEAEALKSDEAEARKSDEAEARKSDEAEARKSDEAEALKSDEAEALKS
DEAEARKSDEAEARKSDEAEALKSDEAEARKSDEAEARKSDEAEARKSDEAEALKS
EGPKGTGGPGSGGSHNKKKSKSMNMLLM"
mat_peptide
734..1591
/product="S antigen"
BASE COUNT 1601 a 270 c 504 g 1463 t
ORIGIN
1 bp upstream of EcoRI site.
Alignment Scores:
Pred. No.: 91.7 Length: 3838
Score: 51.00 Matches: 9
Percent Similarity: 80.00% Conservative: 7
Best Local Similarity: 45.00% Mismatches: 4
Query Match: 39.53% Indels: 0
DB: 3 Gaps: 0
US-09-847-539a-6_COPY_59_86 (1-28) x PFAS47 (1-3838)
QY 8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValysAlaAspAsnAla 27
Db 953 AGTGATGAGCAGGAGCAGCTATAAAGTGATGAGCAGGAGCAGCTATAAAGTGATGAGGCA 1012
RESULT 29
AE005155
LOCUS
DEFINITION
Halobacterium sp. NRC-1 12423 bp DNA linear BCT 12-FEB-2001
complete genome.
ACCESSION
AE005155 AE004438
VERSION
AE005155.1 GI:10584262
KEYWORDS
Halobacterium sp. NRC-1.
SOURCE
Halobacterium sp. NRC-1
ORGANISM
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halobacterium.
REFERENCE
1 (bases 1 to 12423)
Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
Shukla,H.D., Lasky,S.R., Balliga,N., Thorsson,V., Sbrogna,J.,
Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
Isenbarger,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
DasSarma,S.
From the cover: genome sequence of halobacterium species NRC-1
Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)
11016950
2 (bases 1 to 12423)
Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
Shukla,H.D., Lasky,S.R., Balliga,N., Thorsson,V., Sbrogna,J.,
Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
Isenbarger,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
DasSarma,S.
Direct Submission
Submitted (14-JUL-2000) Institute for Systems Biology, 4225
Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES
source
1..12423
Location/Qualifiers
/organism="Halobacterium sp. NRC-1"
/strain="NRC-1"
/db_xref="taxon:64091"
/plasmid="pNRC200"
147..1016
gene
147..1016
/genes="VNG6208C"
147..1016
CDS
147..1016
/genes="VNG6208C"
/note="conserved hypothetical protein"
/codon_start=1

```

```
/protein_id="CAC49491.1"
/db_xref="GI:15140978"
/translation="MCVPEQNGKWAEDARRMVELGSLSVRIGEFAPAKIEPRSGE
FHWLEDAIDVLGKAGLVLTGTPAAPPKWLNYRPEILPVDATGAVRFGARRHY
CFSSRRSEAKARITEAMARYGEHYIYHAWQTDNEYGDHDTIYSYSAEYAGFRLWL
AERYGSDIDELNRYGTFWNRYSDFEEDIDPNNLVEPSPTHGVDFTRFSDDQVKS
NKAODIITRAHSPQRPVTHNPMSQNTDFHYRVGVEDIDIASWDVYPMGGLNGRLAAK
DKHYLRVGDDQDAFNHDLVYAVRGVRVWMEQPGPVNNAHNSPADGCMVRLWTH
LAYAGVDMYSYFWRQAPFAEQFHAGLLLPNSEADQGYLEVAEYVAKRELPEGEV
RCKAQVALVLYESRMYRVLPGQSYSAVAALDMYSYVARLVGDVDFICQHSIDIS
YKLLADPLVTAEEAFVERTRADAKVVGARGSKTRDMHPIPEGLPGPLAKLIDIS
VSRVSEPEFHSYLVGNEAYEAGWRETVRTSETVIAESDFGEYRNGAPALVSDKA
RYLAANGALLDVKVIGDALGNAGLESPLDGLRLVTRRGNLRFAPNFRIPAEVPAG
ASAEYVGGRTLKPDVAIWR"
/complement(5426..6460)
/gene="SMB20987"
/complement(5426..6460)
/gene="SMB20967"
/function="MISCELLANEOUS; Not classified regulator"
/notes="Product confidence : putative"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcriptional regulator, lacI family
protein"
/protein_id="CAC49492.1"
/db_xref="GI:15140979"
/translation="MQRPTKIARTETGLSTATVSKALKHSPQVRPQVTRAVVLOAER
VGYELNHGVOLRTGKYQVAAITGPKENEGVEYLAQLLSGSMWLEESYRVS
LYAVDFEESQETTSQVSMKADGIIISGRADDPRIRLMQEAEPVTVYGMSTHNA
PHAYVDANDEQILSRVARIERHRIALLNPKFOAYGVVRLESYRKALEKAGLPY
DPAVAHGRITPAFGRENVTMSALADPTAYICANEATGALGAFSGFHARGLVHCRDA
VINATDLNVSQYAPPTISYPLIGPESALLGFILRRMEGEPDALQTLMLPDLE
RSDRLRPR"
/complement(6468..8180)
/gene="SMB20968"
/complement(6468..8180)
/gene="SMB20968"
/function="MISCELLANEOUS: Hypothetical/Partial homology"
/notes="Product confidence : hypothetical"
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49493.1"
/db_xref="GI:15140980"
/translation="MLFDIDHVPFSRGRFLTSTMRVPGRDGERALYLRRVSGDER
PSLGRCLRVFLDGEPRPAATRFQSLDQLVARTGEGIVREVGDRHIRGENAGV
RFLVGEGRDYVYRTPGEGCCLVAAGENVKPIPRGLCELVAGSKWDRDRSTVFTL
SGEFTFEGNVDFRALLPPEPGSGFEEAHAAVTAFFGAWYRSIPAGVPGQEAORLAA
YLLWANTVPAEGLVTRPAIYMSKNGMININSDNANFSAAGVADEELAFQCPAAIFD
HQDASGLLPDYNDREALFAFTKPPVHGWAVSCWARENPAELTPERRAYLRDAIGRQV
SYWTHGRAGNVLPVPSYFVHNDSGMDNASFFAEGGPLVSPDLVFLILACBALANLE
DQKARWLRADIAQLVLTGTGETFAARLAADPERILPGESIQFQPLLIGSR
LPAAQCQLVRLADEGAFITDWTGETFAARLAADPERILPGESIQFQPLLIGSR
QEGCALAREIAQKFCALANAHGMAENDFARSGRGLRORAFATSAVYLLLAQSLRDEK
PVKS"
/complement(8184..8996)
/gene="SMB20969"
/complement(8184..8996)
/gene="SMB20969"
/function="CELL PROCESSES: Transport of small molecules;
carbohydrates, organic acids, alcohols"
/notes="Product confidence : putative"
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative sugar uptake ABC transporter permease
protein"
/protein_id="CAC49494.1"
/db_xref="GI:15140981"
```

```
/translation="MRKTVLYATLILLALSFLAPRYTWTFAVKSTAEYLRRPPVLPW
SEWHWENFAAWSKQPFCTYLSNLSLVVVLSTIGOLLSSLVAFGFAFRFPRGRDALF
VLLATMIPMDVDMPIYMEFNMLCNWLTCLPLVPAYFADAFVFLRLQYINTIPM

Alignment Scores:
Pred. No.: 5,72e+03 Length: 303100
Score: 52.00 Matches: 11
Percent Similarity: 77.27% Conservative: 6
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 40.31% Indels: 0
DB: 1 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x RME603646 (1-303100)

QY 4 LeuGLuAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
DB 38610 ATGGATGATCTGCCGACGGAACGACCGATTCGTCGACGCGGTGATCGG 38551
QY 24 AlaAsp 25
DB 38550 CAGGAC 38545

RESULT 27
LOCUS AY061146 2151 bp mRNA linear INV 08-NOV-2001
Drosophila melanogaster LD12115 full length cDNA.
ACCESSION AY061146
VERSION AY061146.1 GI:16768949
KEYWORDS FLI_CDNA.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,
Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
Direct Submission
Submitted (30-OCT-2001) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our Web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
Location/Qualifiers
1..2151
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/map="2288-2288"
1..2151
/gene="CG7289"
/notes="alignment with genomic scaffold AE003584"
/db_xref="FLYBASE:FBgn0031379"
```

MVVYGVVLVDSIKEDAGEDLQVCGTALFPHLLGKAVVSSTT"

# Alignment Scores:

Pred. No.: 5.3e+03 Length: 281993  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 3 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AE003619 (1-281993)

Oy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValVallys 23

Db 246775 CTGGAAGCTCTGTTCAAGCAGGTGAGTCTCTCCAAAGCCGCGAGTTGCTCTTGTGGCA 246834

Oy 24 AlaAspAsn 26

Db 246835 GCCGACAAT 246843

## RESULT 26

LOCUS RME603646/c

DEFINITION Rhizobium meliloti 303100 bp DNA linear BCT 16-AUG-2001  
segment 5/6.

ACCESSION AL603646 AL591985

VERSION AL603646.1 GI:15140973

KEYWORDS

SOURCE Sinorhizobium meliloti.

ORGANISM Sinorhizobium meliloti.

REFERENCE 1 (bases 1 to 303100)

AUTHORS Finan.T.M., Weidner.S., Wong.K., Buhrmester.J., Chain.P., Gouzy.J.,

Vorholter.F.J., Hernandez-Lucas.I., Becker.A., Cowie.A., Gouzy.J.,

Golding.B. and Puhler.A.

From the Cover: The complete sequence of the 1,683-kb pSymb

megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti

Proceedings of the National Academy of Sciences of the United

States of America. 98 (17), 9889-9894 (2001)

11481431

epub ahead of print

2 (bases 1 to 303100)

WEIDNER.S.

Direct Submission

Submitted (07-JUN-2001)

Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld,

Germany

Submitted on behalf of Universitaet Bielefeld, Biologie IV

(Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany and

Department of Biology, McMaster University, 1280 Main Street West,

Hamilton, Ontario, L8S 4K1 Canada

mailto:Stefan.Weidner@Genetik.Uni-Bielefeld.DE

PEXO, pSymb.

## FEATURES

source

gene

complement(167..967)

/gene="thid OR Smb20962"

complement(167..967)

/gene="thid OR Smb20962"

/EC-number="2.7.4.7"

/function="Small Molecule Metabolism; Biosynthesis of

cofactors, carriers; thiamin"

/note="Product confidence : probable

Gene name confidence : putative"

/codon\_start=1

/transl\_table=11

/evidence-not\_experimental

/product="probable phosphomethylpyrimidine kinase protein"

/protein\_id="CAC49487.1"

/db\_xref="GI:15140974"

/translation="MTAIALSIAGSGGGAGIOADLKTFSALGVYCASVITATQON

TRGVAVEDYSAETVSQMDAVESDLVKAIGMVSRETIAIADGLRRCKRAVY

DPVAVTSGDALLRPDAVALIEELLALVYVFNLAALMTRAIAGDAENARQA

EAMTKGAHVAVGGHKGGOEATDFDGDITLVRLPAGRIETNDHGTCTLSAAIA

AGLAKGVPLIEAVSKAAKYLHAAISAADRLIEIGQGRGVPVHFHFWKDD"

1002..1307

/gene="Smb20963"

1002..1307

/gene="Smb20963"

/function="MISCELLANEOUS; Not classified regulator"

/note="Product confidence : putative"

/codon\_start=1

/transl\_table=11

/evidence-not\_experimental

/product="putative transcriptional regulator protein"

/protein\_id="CAC49488.1"

/db\_xref="GI:15140975"

/translation="MRVVRWTSIDLTQTKHGGDPKGDEKANLPLAAIFPFLRRHDP

DOYCGSGSTPSOHLVLPSECAIYAAVACQARAKACKTYLLGQPAKIVPQIGKL"

1412..2071

/gene="Smb20964"

1412..2071

/gene="Smb20964"

/function="CELL PROCESSES; Protection responses"

/note="Product confidence : putative"

Gene name confidence : hypothetical"

/codon\_start=1

/transl\_table=11

/evidence-not\_experimental

/product="putative anti-oxidant protein, AhpCTSA family"

/protein\_id="CAC49489.1"

/db\_xref="GI:15140976"

/translation="MSLRINDTADPTAETQGTINPHEWIGDCAVLFVSHPKNFTPV

CTTELGMAGIEPEFRKRVKIIIGISVDPPVESHKWNKDIKVAATGFEVDYPLIGDRDL

KVAKLYMLPAGAGTSEGRTPADNATVRSYVIGPKKIKLITYPTMTGRNFNEIL

RAIDSIQLTAKHQVATPANWQGEDVITTAASVNEDAVQRFSGFDVLPYLRKTKPT

A"

complement(2189..3187)

/gene="Smb20965"

complement(2189..3187)

/gene="Smb20965"

/function="MISCELLANEOUS"

/note="Product confidence : putative"

Gene name confidence : hypothetical"

/codon\_start=1

/transl\_table=11

/evidence-not\_experimental

/product="putative exported

6-aminohexanoate-dimer hydrolase (EC 3.5.1.46) protein"

/protein\_id="CAC49490.1"

/db\_xref="GI:15140977"

/translation="MLVTLLALIGFHSAAAQDSQSVTTLRLTRELKPLKTVVY

ARDGETAEKGYRGTFTSESTNISKASKISVALVGAIDKGLDGPQKTAFLKED

LPSQADPRINDITIGNLLSMOAGLRSGPNYGRWVSRNRVRFALQAPFDDEPGRM

LYSTAPTHLLSAILTKVGGKPTLALAREWLDPVVEGFRIGAWERDPQGYLGGNQAMS

ARSLLAGELYRNERTADGRQVVPADWVELSWRRPTASRTGQGYGVGFTRIGDE

DVYIANGYGGMLXIVPSLNLTVVMTSDSGPSARNGYRDALHGVLEIISTAKAD"

complement(3296..5236)

/gene="lac22 OR Smb20966"

complement(3296..5236)

/gene="lac22 OR Smb20966"

/EC-number="3.2.1.23"

/function="Small Molecule Metabolism; Degradation; carbon

compounds"

/note="Product confidence : putative"

Gene name confidence : putative"

/codon\_start=1

/transl\_table=11

/evidence-not\_experimental

/product="putative beta-galactosidase protein"



REFERENCE 2 (bases 1 to 281993)  
ADAMS,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.  
TITLE Direct Submission  
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
COMMENT On Oct 9, 2000 this sequence version replaced gi:7297302.  
FEATURES  
source  
1. .281993  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
mRNA  
join(421..927,1513..1971,2323..2899,2959..3222,3299..3351,  
3451..4223)  
/gene="BCDNA:LD21794"  
/product="CT21997"  
/db\_xref="FLYBASE:FBan0007115"  
/db\_xref="FLYBASE:FBgn0027515"  
gene  
<421..>4223  
/gene="BCDNA:LD21794"  
/note="CG7115"  
/map="28D1-28D2"  
/db\_xref="FLYBASE:FBan0007115"  
/db\_xref="FLYBASE:FBgn0027515"  
join(437..519,1513..1971,2323..2899,2959..3222,3299..3351,  
3451..4223)  
/gene="BCDNA:LD21794"  
/product="CT21999"  
/db\_xref="FLYBASE:FBan0007115"  
/db\_xref="FLYBASE:FBgn0027515"  
join(1595..1971,2323..2899,2959..3222,3299..3351,  
3451..3754)  
/gene="BCDNA:LD21794"  
/note="BCDNA:LD21794 gene product [alt 1]"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0007115"  
/db\_xref="FLYBASE:FBgn0027515"  
/protein\_id="AAF52564.1"  
/db\_xref="GI:7297303"  
/translation="WDELEDKVFYQTYVSHMKILSKFVGPSSINSPLAYIWKLCRL  
YVLEPFIKGLFVLLYLVQVDVMSRSLIGRIQATIGRQKAVKMMELSSASGDHQ  
SWEKMQSSAFVILGRPRMEDRFIIENINNTGISFFAVFDGGHGEFAADFAKDV  
LYKNYIKIENSKLLTEGNSGDYDKSPYLARKOSRKDKANKENTEPTAGVMRKDSL  
KAHSTTADCSAIKQKTEASIADIVTVOLNSAMRASGNLGAAKDSFLNNNNGONGAA  
NAPPNYEAKCYIEHGRNFGKLIITDEIMSDYKLVEQAKRATINAGTTALLAIVOGS  
KLIVANVGDSSRGVMDWGIAIPLSFDPKPOQVREKRIRHDAGGFIAPRGVWRVAGVL  
ATSRALGDTPLDKNLVIATPDLITFELNDHRPHLILASDGLMDTFFSNEEACTFALE  
HLKPEDFGAKSAMESYKRGSDNITVLVIVFKNDYKIKIGSSAGKAGEESLKVPAKSQ  
PVAPAVQRSNSIKTK"  
CDS  
join(1595..1971,2323..2899,2959..3222,3299..3351,  
3451..3754)  
/gene="BCDNA:LD21794"  
/note="BCDNA:LD21794 gene product [alt 2]"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0007115"  
/db\_xref="FLYBASE:FBgn0027515"  
/protein\_id="AAF52565.1"  
/db\_xref="GI:7297304"  
/translation="WDELEDKVFYQTYVSHMKILSKFVGPSSINSPLAYIWKLCRL  
YVLEPFIKGLFVLLYLVQVDVMSRSLIGRIQATIGRQKAVKMMELSSASGDHQ  
SWEKMQSSAFVILGRPRMEDRFIIENINNTGISFFAVFDGGHGEFAADFAKDV  
LYKNYIKIENSKLLTEGNSGDYDKSPYLARKOSRKDKANKENTEPTAGVMRKDSL  
KAHSTTADCSAIKQKTEASIADIVTVOLNSAMRASGNLGAAKDSFLNNNNGONGAA  
NAPPNYEAKCYIEHGRNFGKLIITDEIMSDYKLVEQAKRATINAGTTALLAIVOGS  
KLIVANVGDSSRGVMDWGIAIPLSFDPKPOQVREKRIRHDAGGFIAPRGVWRVAGVL  
ATSRALGDTPLDKNLVIATPDLITFELNDHRPHLILASDGLMDTFFSNEEACTFALE  
HLKPEDFGAKSAMESYKRGSDNITVLVIVFKNDYKIKIGSSAGKAGEESLKVPAKSQ  
PVAPAVQRSNSIKTK"  
complement(join(4497..5207,5641..5852,5932..6106))  
/gene="Rack1"  
/product="CT21991"  
mRNA  
/db\_xref="FLYBASE:FBan0007111"  
/db\_xref="FLYBASE:FBgn0020618"  
complement(<4497..>6106)

/gene="Rack1"  
/note="CG7111"  
/map="28D1-28D2"  
/db\_xref="FLYBASE:FBan0007111"  
/db\_xref="FLYBASE:FBgn0020618"  
complement(join(4572..5207,5641..5852,5932..6040))  
/gene="Rack1"  
/note="Rack1 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0007111"  
/db\_xref="FLYBASE:FBgn0020618"  
/protein\_id="AAF52566.1"  
/db\_xref="GI:7297305"  
/translation="MSETLQLRGLTHNGWVQIATNPDKDPTIISASRDKTLIVMK  
LTPDEDNYGYQKRLYGYSHFISDVVLSDDCNVYALSGSWDOTLRWDLAAGKTTBRF  
EGHTKDVLSVAFSADNRQIVSGSRDKTIRKLNWLAECKFTIQDGDHWDVSCVRFSPN  
HSNPITVSCGWDRTVKVNLANKLNHHNGYLVNTVTVSPDGLSCTSGGKDSKAL  
LWDLNDGNLYTLTEHNDIIINALCFSPNRYLVCVAYGPSIKIMDLACKKTVTELRPEVY  
SPTSKADQPCLSLAWSOTGOTLFGAGYSNTIRVQVSVSAH"  
join(6617..7024,9318..9893,10340..10530,10605..11524,  
11707..11716)  
/gene="mts"  
/note="Nucleotide sequence of the Celera sequence differs  
from the published sequence for this transcript."  
/product="CT21977"  
/db\_xref="FLYBASE:FBan0007109"  
/db\_xref="FLYBASE:FBgn0004177"  
<6617..>11716  
/gene="mts"  
/note="CG7109"  
/map="28D1-28D3"  
/db\_xref="FLYBASE:FBan0007109"  
/db\_xref="FLYBASE:FBgn0004177"  
join(6923..7024,9318..9893,10340..10530,10605..10677)  
/gene="mts"  
/note="mts gene product; Nucleotide sequence of the Celera  
sequence differs from the published sequence for this  
transcript"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0007109"  
/db\_xref="FLYBASE:FBgn0004177"  
/protein\_id="AAF52567.1"  
/db\_xref="GI:7297306"  
/translation="MEDKATTKDLDQWIEQLNECNLTQVOTLCKDSLQAKKELL  
SRESNVQKCPVTVCGDVHGFHDLMEPLRIGGKSPDTNYPMGDYVDRGYYSVERV  
TLVALUKVIRERITILRGNHESRQITVGFDECLRTKGNANWKYIFDLFDYLPFL  
TALVDQIIFLHGGLSPISDLSLDHIKALDRQLQVPEHGPCMDLLSDPDRGGMGISP  
RGAGYTFGDISFTFNNTNGLTIVSRAHQLVMEGYNWCHDRNVYVTFISAPNYCYRCGNQ  
AALMELDLSLKFSLQFDPAPRRGEPHVRTRTPDYEL"  
join(19410..>20021  
/gene="CG14537"  
/product="CT34267"  
/db\_xref="FLYBASE:FBan0014537"  
/db\_xref="FLYBASE:FBgn0031954"  
/evidence=not\_experimental  
<19410..>20021  
/gene="CG14537"  
/map="28D2-28D3"  
/db\_xref="FLYBASE:FBan0014537"  
/db\_xref="FLYBASE:FBgn0031954"  
/evidence=not\_experimental  
19410..20021  
/gene="CG14537"  
/note="CG14537 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0014537"  
/db\_xref="FLYBASE:FBgn0031954"  
/evidence=not\_experimental  
/protein\_id="AAF52568.1"  
/db\_xref="GI:7297307"  
/translation="MSEPLSGRWLRVFNNGLLLASAVCTGRKMLPEHPFAMVACVVV  
GFSAVFLLRVIFASQPECCQLRDTISVLEPLANFDPMKIDVSTGSLSLIALGH  
APFVLPFLFCGLGSLAKRRDCAFDSLRNLTVLGNITVSLGLAYVFNFLYLRMLV



```

QY      24 AlaSpasn 26      169427 bp      DNA      linear      INV 07-JUN-2001
LOCUS   Drosophila melanogaster, chromosome 2L, region 28B-28B, BAC clone
DEFINITION   BACR08101, complete sequence.
ACCESSION   AC007416
VERSION     AC007416.12 GI:14327741
KEYWORDS    HTG.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 169427)
AUTHORS     Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brannon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,Y., Doup,L.E., Doyle,C., Garg,N.S., George,R.A.,
            Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., Hostin,D., Howland,T.J.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F.,
            Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
            Sequencing of Drosophila chromosome 2L, region 28B-28B
            Unpublished
REFERENCE   2 (bases 1 to 169427)
AUTHORS     Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
            Direct Submission
            Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
            On Jun 7, 2001 this sequence version replaced gi:7143379.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdg@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source            1..169427
                        /organism="Drosophila melanogaster"
                        /strain="y; cn bw sp"
                        /db_xref="taxon:7227"
                        /chromosome="2L"
                        /map="28B-28B"
                        /clone="BACR08101 (D605)"
                        /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                        Drosophila melanogaster BAC library, partial EcoRI in
                        pBACE3.6)"
                        /note="Alternate"
BASE COUNT           51061 a 35853 c 34873 g 47640 t
ORIGIN

```

---

```

Alignment Scores:
Pred. NO.:      3.13e+03      Length:      169427
Score:          52.00         Matches:    11
Percent Similarity: 65.22%      Conservative: 4
Best Local Similarity: 47.83%      Mismatches: 8
Query Match:    40.31%         Indels:     0
DB:             3            Gaps:         0

US-09-847-539A-6_COPY_59_86 (1-28) x AC007416 (1-169427)

QY      4 LeuGluaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
LOCUS   Drosophila melanogaster genomic scaffold 142000013386055 section 12
DEFINITION   of 63, complete sequence.
ACCESSION   AE003619
VERSION     AE003619.2 GI:10728652
KEYWORDS    HTG.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 281993)
AUTHORS     Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Heit,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Bernan,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dunn,P., Dupis,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
            Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskaas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
            The genome sequence of Drosophila melanogaster
            Science 287 (5461), 2185-2195 (2000)
            20196006
TITLE
JOURNAL
MEDLINE

```

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 36993)  
AUTHORS Adams, M. and Venter, J.C.  
JOURNAL Direct Submission  
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
COMMENT This sequence was identified as CDW:10212995 by the submitter.  
For more information on this record e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
1. .36993  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 9781 a 8675 c 8725 g 9812 t  
ORIGIN  
|||||  
Alignment Scores:  
Pred. No.: 649 Length: 36993  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC018007 (1-36993)  
QY 4 LeuGluaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLys 23  
Db 30206 CTGGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAGCGCGCAGTTCCTCTCTGTGGCA 30265  
|||||  
QY 24 AlaAspAsn 26  
Db 30266 GCCGACAAT 30274  
|||||

RESULT 23  
AC011910  
LOCUS  
DEFINITION Drosophila melanogaster clone RPC198-2017, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 39 unordered pieces.  
AC011910.1 GI:6056126  
VERSION  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 133155)  
Muzny, D.M., Adams, C., Bailey, M., Barberia, J., Blankenburg, K.,  
Bodota, B., Bock, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,  
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,  
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,  
Dugan-Rocha, S., Durbin, K., Fernandez, C., Ferraguto, D.,  
Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,  
Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hogue, M.,  
Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,  
Kelly, S., Kondejowski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,  
Lichtarge, O., Liu, J., Liu, W., Logan, O., Lu, J., Lucier, R.,  
Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S.,  
Nash, S., Nelson, A., Nguyen, K., Nguyen, N., Nguyen, S., Oswal, G.,  
Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M.,  
Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, B.,  
Shen, H., Simon, M., Sparks, A., Stamps, A., Suckang, R., Taber, P.,  
Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M., Watlington, S.,  
Weinstock, G., Weinstock, I.R., Williamson, A., Worley, K., Wren, J.,  
Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.

TITLE  
JOURNAL  
REFERENCE 2 (bases 1 to 133155)

AUTHORS  
TITLE  
JOURNAL

Worley, K.C.

Direct Submission

Submitted (16-OCT-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 39 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 854: contig of 854 bp in length  
\* 855 1695: contig of 841 bp in length  
\* 1696 2535: contig of 840 bp in length  
\* 2536 3358: contig of 823 bp in length  
\* 3359 4171: contig of 813 bp in length  
\* 4172 5125: contig of 954 bp in length  
\* 5126 6105: contig of 980 bp in length  
\* 6106 6952: contig of 847 bp in length  
\* 6953 7603: contig of 651 bp in length  
\* 7604 8573: contig of 970 bp in length  
\* 8574 9724: contig of 1151 bp in length  
\* 9725 12519: contig of 1247 bp in length  
\* 10972 12519: contig of 1548 bp in length  
\* 12520 14371: contig of 1852 bp in length  
\* 14372 15264: contig of 893 bp in length  
\* 15265 16905: contig of 1641 bp in length  
\* 16906 19536: contig of 2631 bp in length  
\* 19537 21604: contig of 2068 bp in length  
\* 21605 22985: contig of 1381 bp in length  
\* 22986 24363: contig of 1378 bp in length  
\* 24364 26424: contig of 2061 bp in length  
\* 26425 28204: contig of 1780 bp in length  
\* 28205 31215: contig of 3011 bp in length  
\* 31216 33382: contig of 2167 bp in length  
\* 33383 36316: contig of 2934 bp in length  
\* 36317 40441: contig of 4125 bp in length  
\* 40442 44615: contig of 4174 bp in length  
\* 44616 49292: contig of 4677 bp in length  
\* 49293 53647: contig of 4355 bp in length  
\* 53648 59590: contig of 5943 bp in length  
\* 59591 61868: contig of 2278 bp in length  
\* 61869 67152: contig of 5284 bp in length  
\* 67153 71513: contig of 4360 bp in length  
\* 71513 76928: contig of 5416 bp in length  
\* 76929 82257: contig of 5329 bp in length  
\* 82258 89419: contig of 7161 bp in length  
\* 89419 103318: contig of 13900 bp in length  
\* 103319 113206: contig of 9888 bp in length  
\* 113207 133155: contig of 19949 bp in length.

## FEATURES

Location/Qualifiers

source

1. .133155

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="RPC198-2017"

BASE COUNT 38122 a 28987 c 27853 g 38141 t 52 others

## ORIGIN

## Alignment Scores:

Pred. No.: 2,44e+03 Length: 133155  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC011910 (1-133155)

QY 4 LeuGluaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLys 23

|||||

Db 99660 CTGGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAGCGCGCAGTTCCTCTCTGTGGCA 99719

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (18), 10445-10450 (2001)  
 PUBMED 11517334  
 REFERENCE 2 (bases 1 to 17595)  
 AUTHORS Richards, S., Hekmat-Scafe, D.S., Hurd, D.D., Caprette, D.R., Saxon, W.M., Carlson, J.R. and Stern, M.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-OCT-1998) Biochemistry and Cell Biology, Mail Stop 140, William Marsh Rice University, 6100 South Main, Houston, TX 77005, USA

FEATURES  
 source Location/Qualifiers  
 1..17595 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /chromosome="2"  
 /map="28E1"  
 360 /note="push3420 P element insertion site"  
 repeat\_region  
 mRNA  
 join(370)..513,695..789,913..1325,1581..3747,3808..4335,4404..10389,10447..10651,10701..11393,11455..13118,13175..13657,13715..15650,15706..17302,17366..17584)  
 /gene="push"  
 /product="pushover"  
 370..17584  
 /gene="push"  
 /gene="push"  
 join(395)..513,695..789,913..1325,1581..3747,3808..4335,4404..10389,10447..10651,10701..11393,11455..13118,13175..13657,15706..15715)  
 /gene="push"  
 /codon\_start=1  
 /product="pushover"  
 /protein\_id="AAD20449.1"  
 /db\_xref="GI:4426610"

variation /gene="push"  
 /replace="t"  
 3600  
 variation /gene="push"  
 /replace="t"  
 17558..17563  
 /gene="push"  
 17584  
 /gene="push"  
 4795 a 4143 c 4251 g 4406 t  
 BASE COUNT  
 ORIGIN

Alignment Scores:  
 Pred. No.: 301 Length: 17595  
 Score: 52.00 Matches: 11  
 Percent Similarity: 65.22% Conservative: 4  
 Best Local Similarity: 47.83% Mismatches: 8  
 Query Match: 40.31% Indels: 0  
 DB: 3 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AF096896 (1-17595)

Qy 4 LeuGluaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValVallys 23  
 |||||  
 Db 7027 CTGAAGCTGTTCAAGCAGGTGAGTCTCTCAAGCCGCGAGTTGCTCTCTGTGGCA 6968  
 |||||  
 Qy 24 AlaAspAsn 26  
 |||||  
 Db 6967 GCCGACAAT 6959

RESULT 22  
 AC018007  
 LOCUS  
 DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
 ACCESSION AC018007  
 VERSION AC018007.1 GI:6553183  
 KEYWORDS HTG; HTGS\_PHASE2.  
 SOURCE fruit fly  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

FEESNFTQVTAALAAADKLMQITCOTICOLHDATAVLIRFTIYLRPVSYVE  
 TKNLGALMLCGERCPASASSMDYNAVNLKCKHPSTTKSIMPSSSSGSG  
 ASNDKPKSEIKRSDLSVILQOLIAPLRPGKWTWPLSEVDTCTQILAAVVE  
 YPQNGVDLLDVCVPLNIRKSMETTINGKSLYLPLTQVEATVATVNSHML  
 TDTLSQAALLIEMOPLNSRDIEMGCIAGLYNAVLTSIATVILGMSQAASSQKQ  
 TASTSQSGVSSGGGOSNRHDDFEDQACSIVNKALEIYSNIGHMFKTSARIRHVN  
 HLCYGSWLLSGIQAGSGSGSDSKASKATKSGSEAGTAPTPPIARVNL  
 KVQOFGELAAIANHSIKLSLIEDLVKVAACGQSLESTELPEPAQDFIYSSSL  
 ERIVRLNTLHQLTFLAVAYRKAETLKRASAKDRTECEPISYSDSTYENLSL  
 CSNSEDDESILGHWFETLSPETHDDNANTSTQERAKQSAVLPKDEPHYLDL  
 AADLFCFLDGLANRHMORYKAGVSDQOOLMLMANIKDFDVRNRETDOGSNA  
 PAASAGTASAGTSKWTSMIRFSGAAGRYIHNLTSLLSEQLQNLQHLHSISW  
 STNTNTPVAGYPTSLVQIILLKPTQKEACLSVWHRLINTLVEGVCSNSTAD  
 SYEDLNIHQAQLLFHSLNLMQKSIULLTAGVIRCAEVCGRISDRPVKNSDI  
 MLLSRLLFLVLMKLYNAPPELLDOVRNLFSSVSMPTQKITDILNCRKLNSYC  
 RHDIENFRKSAGEYSSIRPTIYSLVWGDPTSYWAQEPKLDGLANFILCIPDKLK  
 YPLVDALDILSTDSMSYKSKDKKASMHNLCAIQCTFIANKLNLGUPPSTSHVE  
 SLKAERSPNLHLSMISIRLPLASHYLVSSLIKQGYTQYAEITLHVGDIGADIKY  
 SKQITILGVFAFNSQNGTPTRLSDILFDSLVAHQAVANKEGLKWPKESEDA  
 AGEQACTSLTSSNDPELYSSNESIDQKLKDDGKLSLDQKYNOLYNELMVKLM  
 LKLVYHVMGCVYKMLKEAADLRQPAAGSGGAPLQDRAMFNESMLRFLVKTMMX  
 LEKPVAYAMQROFFKQSGSLTLLSFTGTSIPVYARKMLQFVNRMLFQLSQADS  
 QFHEDLVECFSELATVDVARLKOWLGHIIYGPVNTDVTSEALDTCRLMTSLQ  
 SSSSSNAOQTPNMAYSAMPSTSDQDPMIEYDCCTAAGGSPNCTGGAANTSO  
 ILSLWAOAPDPPESSQACDHSERNGALLSFYKSLVKDQSKASQIAPPLFOALL  
 QLGQTLSPQEGCFADVLQIMITLADASPARGHVALFNTILWELAKLQLPDKHL  
 KHAENVAQURLYSELLQISGRSQHNPWDELDQTDIDELYDELAEEGQDLSLD  
 DSDEDTLNKLCTFSQTKFMQHWYHCHTCNNMTVGVCSVCARCKHGHVSYAK  
 YGNPCDCGAKDCGSCQALSRIGSNEVRDSAGISYLPSSHMSLACKKRSLLPVGO  
 VLPKRSLSLTERIAVLTKLEPYRETIQHODQMLLVVRCILEYDILLPTINENCLY  
 SVLCCHKRATAALERHLLEQSFQVTLQIMPATLGSOGAFENVRMNYSGDQGTIKH  
 LLSGVRVRRPACCLSPHGRROQLAVSHEKGVITLIQLSALLQKQADAKRKLTLTOL  
 SSAPDIACITVLSLANPCNEDCLAVGLKECHILTFSSSGSTNEHVYNPQLENGNYIK  
 KAVMLQSQTLLAVYSDVYLYDLAVDTYSPKYLYVAVGKTRDCFTFYQDQGNML  
 TFASGGIYITQOLDOQSILVHGDFTVNTLELSHOIKDIAGHIGGGGVSTIYSHTLQ  
 LLFYSYSGSGSPFLNVSQGVGIYHLDTNASKSKSGPLQPLQVSTVEVTHGPL  
 VYASMQTSNPNILMITPFIYQETKAQSAKSRIMDVGVIRHVAAGVETKTLLLICE

ASNDKSPKSEIKRSDLSVLQOOLIAPIEPGKNTWVPLSEEVTDCTEQILAAANVE  
YFOEQNGVTLDDVCVSLPILNRYRSKYMETINGCKSLYLPLTOVEATVAKSMNHML  
TDLTILSOAALIEQOPLTPSRIEBSLWCGTALYNNAVLTSLATISVSKSASSQKQ  
TASTSGSGSGSGSGGKNDHDDFEDQACSIVNKALEIYSNIGHMFMTSAKHVYQN  
HUCYGSWLLSIGIAGMAGSGSGSSDSKASAKATKSGSEAGTAPPTPIARVNL  
KVGQFGLNAGIANHSTIKLSLEITDLKVEAACQSLSELETPAEPADILQNTSSL  
ERIVRVNATLTHLOFTLAVYAVKAKTILKRAKADTECEPTISYSDSTYFNDSLS  
CDSNDESDSSEYLGHWFKEITLSPEHODNANTSQERAEOKSALVPLDRPEYLDL  
AADLFCFLDOLANRHAYQRYVAVAGVSDQOQLMANNILKDFEDRVNRENDQSGSNA  
PASAGAGTSAGATSWQTSIMIRFSGAAGRYTHNLSISLSEQLQSLNQLHLSLSPW  
STDNTPWQVYPTSLVSLVQLLPLTPQKEAKELSVYHRLINTLVGECVCSNSTASD  
SYEDNLTEHAQLLFLFHLNLMQKKSILLITAGGVIRCAEVCKI9EDRPRVKNQSI  
MLLSRLLLEFVNLKHLNAPPELLDQVRNLFVSVMPTDKITDILNCRKNSIYC  
RHDIGENFRKSGAGVSSIRPTEYSLVMDPEISYMAOEFKLDGLAWNELICTPDKLK  
YPLVADALTILSTIDMSYKSEKDKKASMINLCAIOYCFKIAWKLNLGLPPSTSHVE  
SLKASPNLSHMSVIRPLASHSYLVVSLIKQOYATLWTHVGDIGADIKY  
SLKOTILGVEAFNSOMNGNTPRISDLILFDLSVAHQAVAMANKREGLYKQVNRKESEDA  
AQEGAGTSLTSSSDNPELYSNESIDQKLQDDGKLSLDQKYNQVNLVNLAYKLM  
NSYOLLSVIVGOMLQKLSSTPKALNLIPIVIVSDKPAIMLELHAAPFLKLLPNEDKO  
LIANEPKCLAVNDFAFGKOHPEPYILNVIDAHIIELTRGSNYSTLHTLKHCILKSI  
LOLEPLLPHETANADEVOLKOLLISMLDNRDYLOHSEHREILSLGLTQBAQK  
LLELVHMYGICRMKEFAADLRQPAAGPSGAPLQDQRAFMNESMLFAVLKTFKIM  
LEKPVAAVQKOFFKDRSGSLTLLLSLTSTLSPVYARKMLQVNRFLQSLQSDS  
QFQHDLYECSELATVDVARKLQWHLIIYCPNVSTDTSTSEALDTCRLMTSLTOP  
SSSSSNAQPTNMAVSAMPSISDQDPMETEDCGTAAGGESNPGTGGAAANTSQ  
ILSLWQAAQPPDSESSQADHSDSERGALLISPVKSLVKDQSKASOIAPLFOALL  
OLQOPLISPROGCDADVLOIMITLADASPARGHALPNTLLALELAKLOPKHL  
KHAENVSAQLAYLSLLOSIGFRGSRQINFPWDELQDIDELIDELABEGQDSLLD  
DSDEDLNKKLCTQKQEFNMHWCHTCNMINTVGVCSVCARVCKGHKHDVSYAK  
YGNFCDGKADGSCQALSRIGNSNEVRDAGISYLPSHMSLLAGKRSLSPVQK  
VLTRGKNTIARVILTRKLEPQVRETQQLVLRVCIILEYFDLLPLSNINENCMILY  
STVGCKRTAALERHLLEQSFQVTDQMEATLGSQBGFENVRMNTSGDQGTIKH  
LLSGVVRVAFCLSSPHGRQOLAVSHEKQVTLIOLLSALLQADASKRLTLITOL  
SSAPACTYVISAANPCNEDCLAVGLECHILTFSSSGSTNEHIVVNPQLENGTIR  
KAVPLVSGTILAAVNTSDIYDVIYDVIYSPKYYILVAVGKIRDTYODGNMIL  
TFASGYTYTQLOQSLAVHGFYVNTLELSHQIKIDIAHIGGGVSIYYSHTLQ  
LLFYSGSGRSPFSLTNVSEGVGIYHLDNTSASKSKAGPLOPLQVOWTEVTHPGL  
VYASMTSNPFIILMTPIERYLOEIKQASAKSRIMDVVGI9RHAVAGVEKTTLLILCE  
DGLSIFAQPEYTFMSPQVQPNGLYSLTLMAKGGSGSTSKSNYASGKWT  
RKASQOQKPTAGGQVFPIDFEHCLNADVEFGNLLQIYNKQKRLTFSTGMF  
VASTRNGTEVINNDPNVIVIGRILVIGTQVQVAPQSVITLGTPTPVRRARWF  
DIPTREMIASDILLKVVFAKAPDEHVITLDCIEVYKSKELVGMDESDVTVPG  
SSAPAVSSQSSANFGGFCNICTQIDRMANHLEVMBCALHLGSGVPASQKRAVK  
TASALLLTPNPVOTOARYVLATYGTALYHNYKDGVLQVNFVRELSQMPKLEKL  
ETLREIDPEAYRLVLAVRGITANARPOSALKICVENNDIIVPTLNGIVLELHKVPTL  
DEPNVIRARGCQPTIIVCHLVEIMYGFALADPGQVQGRMTYFIDLLKHDSVISHSA  
KEALTLLSPKMRKRVKVAIVTPPACSTPTSTSTMOLOAAASSASDIIEEAQVY  
DGSVGGGLPEPNADGAAGVGGVGGQOQMLNLEAFMGGFPRLGLPEADDAEAIM  
DIATLSLQOHHGDANALQOGLANLOQIQAOTAMAAVNAANAAVSLGSGDDDEGS  
NVATDGTSLRTSPAEPAGSGSGSGSVESIGTTSARSSNFGDHANASPPROGSKD  
DQEQPGSGVAGSGVAVLSAMSSSEDNEANEDDKLSKLHLDIRITVLESITQHLGTFD  
LCNGLQAIPLQVILMLITDNLNDRQOQVLLHDLTALVLGLCGNKKKGAARMETKC  
PGNEVRLALLSLFVGLMGKTSKGTGTTSPPHQFKDSSFSVASTNYSKSGAEVYA  
LEAINTLVHKNVIGDPYAAAGGLASQASAGSGAGVQOLLKPIKHGKPDLSIL  
IPHNLYKNYPOIFESYDGLLTEIIVRLPYQILRSLSAHPNDYDGCFAMTPTCEYM  
MLNLTILNRQVRKLLMYICGSKFRMYRQCHSDAHPRVVYKRVNIVSSKFTGAPYN  
ANPNMUSDALVELTLEHRTQCBISQMETGNWQKFCVVEDALAMLEIACYQLDDGV  
SPITLIQLOAVCNLPPSGSKQAPQSTSSAGKRLKTRDREKSDTDAYASRDFPAQ  
CGTFVHQIRVACDAILIRFVRIFFLENNITQLRWQAHFSMTGLFEHANEQRKLLN  
IFWNPVLTPTVYGRGTVQVLLGYTLNSTRITERLPEFVSRAVDVLRQTTMYLKL  
FNAPYITLSTLESTLOVNGYILESEPCLVGNPVPWANIKLPSVKSQSVYTTMITYKL  
VQCHTISKILVRIADLKRTKVRTINVYNRPSVOAVELKRPALMKHARSVLSQA  
VOTELKPLPIITACNLMIERADFEETVSGSENIQCPRCSAAVPAYPGVCGNGENV  
FOCHKRANIDKEDPFLCHSGCPKYAKDFSMYARVCCADVPIESAEDRVKTVSLI  
HSSLERADNRTROLTNKOMIELLTKQVAHRSDRVEDNMAVSHYSTSVQNKIITQLL  
AQKYVESRTEFSLEKSVQKACRSELVAYDROQOQPPVPNPGSTGTAENNTNRC  
YGCALASTEOCLTLRAMAYNYDCRVCLYSOGLSELAELHNRCTPILOEVRNLLY  
VLTKDNAEACHMILLOLVTRYKNLWMSIPLISLEAAHQEWLTLVILQDDICWY  
KLKVIFELISNCRLPQPVTAHLPCRLRMONLLICPVLPGSKPNQKVATVTDICSMKM  
FEGVTVQVAKNRDRNHEIYAAWSKPMSPNQAKLNKADQNVASRFGQADAPKSRRE  
VRVAFVLSKYGKRWREYDKQVRIKPLVFNAKIQOPLFNANSRFGQLASLSSSL  
SRTNRRQOALNMLTSLFKHVGGEAGEAEYLMLYKNMATQPMQVYLVLKVLQSIS  
QLLAIEISKVHRMBEYSLSDLSGLYALRQYVELLWLLLECPNIRRTYKTRMLPVLE

SYLALRSLVVORTLIDDQAQEKLEMLEMDMTSGTBEETRAFMEILLIDTVEKTRMNDIK  
TPVVFPERLYSIIHPEEHDSEFYMTLEKQPOEDFLOGRMGLGNPVPSPSEMGIGPLMR  
YMRGLLIDATEEFVETLNNKSQEQVDTEOLYRMANVLDACNGLRMLERISQSLORIS  
NRRELITQLLKFLICVKVRCQEVLCQPEIAGINTLLKVLQMCLOSENDSIQSAVTE  
QLLEIMETILSKAASDTLDSLFQSLTFGGPEYVSAISCTDCPNRNPNSVLRHLIR  
VLAALVYGNFVKMALLCRHKDTLNFNRPDTPDEEFKLELPCVLTNQIGHNCTGG  
TLKDYIVSLGIVERSLAVITERHAPCVKPTLLRDTSDDELKEFISRSPLKYILRLPTGLS  
NHEATQVAISKDIIPIIHRLEQVSDSDHVGSLAENLLEALSTDSTAARVQOVRDPT  
RAEKRRLAHATREKQDLAGMRTNEKQVTAAGSLQKLEKURDETGLTCTICREGYA  
COPDXVLGIYTTKRCNVBEFELKSKTIGYTTVFNHVVHVDCHTSAIRLTRGRDEW  
ERASIQNANTRCNGILLPLWGPVAGPAFAFSACMTRHSSYMQESTQCDISYTSVSHDLK  
LLLVFWANERSGHDHAGGQPSQNMHFVPYLLFYSVYLLLSRASAAROSKTLTLTYLOA  
PPSEKLEGGVEYDCQPLFWANITLSLSHSRELANKKLAHLKRMIAVAGORHVSAPVLC  
KALLAPDRQKVDYTVYKFFLMMNALVDLILNFKTWTSTPKEEDWPTLSLFDYLRKND  
EALLKSTDSILOTLTTEELPCTFSVFEFCQDVGILLHLIEHPDNFIEEILAAALPSTSSN  
"  
variation /gene="push"  
/replace="t"  
2672 /gene="push"  
/replace="t"  
5411.5599 /gene="push"  
/note="putative"  
/bound\_moiety="zinc"  
11402.111542 /gene="push"  
/note="putative"  
/bound\_moiety="zinc"  
15995.16135 /gene="push"  
16130 /gene="push"  
BASE COUNT 4288 a 3943 c 4035 g 3869 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 275 Length: 16135  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 3 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AF096897 (1-16135)  
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
||||| ||||| ||| : : : : : ||||| : : : : : |||||  
Db 5971 CTGGAAGCTCTGTTCAGCAGGTGGAGTCTCTCAAGCGCGGACGTTGCTCTCTTTGGCA 5912  
Qy 24 AlaAspAsn 26  
||||| |||||  
Db 5911 GCCGACAAT 5903  
RESULT 21  
AF096896/7c  
LOCUS AF096896  
DEFINITION Drosophila melanogaster pushover (push) gene, complete cds.  
ACCESSION AF096896  
VERSION AF096896.1 GI:426608  
KEYWORDS fruit fly.  
SOURCE Drosophila melanogaster  
ORGANISM Drosophila  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 17595)  
Yager, J., Richards, S., Hekmat-Scate, D.S., Hurd, D.D., Sundaresan, V.,  
Caprette, D.R., Saxton, W.M., Carlson, J.R. and Stern, M.  
Control of Drosophila perineurial glial growth by interacting  
neurotransmitter-mediated signaling pathways



\* 89805 89904: gap of unknown length  
 \* 89905 95932: contig of 6028 bp in length  
 \* 95933 96032: gap of unknown length  
 \* 96033 102331: contig of 6499 bp in length  
 \* 102332 102631: gap of unknown length  
 \* 102632 111305: contig of 8674 bp in length  
 \* 111306 111405: gap of unknown length  
 \* 111406 120462: contig of 9057 bp in length  
 \* 120463 120562: gap of unknown length  
 \* 120563 130442: contig of 9680 bp in length  
 \* 130443 130342: gap of unknown length  
 \* 130343 140588: contig of 10246 bp in length  
 \* 140589 140688: gap of unknown length  
 \* 140689 149261: contig of 8573 bp in length  
 \* 149262 149361: gap of unknown length  
 \* 149362 162349: contig of 12988 bp in length  
 \* 162350 162449: gap of unknown length  
 \* 162450 173772: contig of 11323 bp in length.

## FEATURES

source

1. .173772  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="RPC1-11"  
 /clone="Rp11-724H13"  
 1. .1282

misc\_feature /note="assembly\_name:Contig17"  
 misc\_feature /note="assembly\_name:Contig20"  
 misc\_feature /note="assembly\_name:Contig22"  
 misc\_feature /note="assembly\_name:Contig23"  
 misc\_feature /note="assembly\_name:Contig24"  
 misc\_feature /note="assembly\_name:Contig25"  
 misc\_feature /note="assembly\_name:Contig26"  
 misc\_feature /note="assembly\_name:Contig27"  
 misc\_feature /note="assembly\_name:Contig28"  
 misc\_feature /note="assembly\_name:Contig29"  
 misc\_feature /note="assembly\_name:Contig30"  
 misc\_feature /note="assembly\_name:Contig31"  
 misc\_feature /note="assembly\_name:Contig32"  
 misc\_feature /note="assembly\_name:Contig33"  
 misc\_feature /note="assembly\_name:Contig34"  
 misc\_feature /note="assembly\_name:Contig35"  
 misc\_feature /note="assembly\_name:Contig36"  
 misc\_feature /note="assembly\_name:Contig37"  
 misc\_feature /note="assembly\_name:Contig38"  
 misc\_feature /note="assembly\_name:Contig39"  
 misc\_feature /note="assembly\_name:Contig40"  
 misc\_feature /note="assembly\_name:Contig41"  
 misc\_feature /note="assembly\_name:Contig42"  
 misc\_feature /note="assembly\_name:Contig43"

misc\_feature /note="assembly\_name:Contig44"  
 misc\_feature /note="assembly\_name:Contig45"  
 misc\_feature /note="assembly\_name:Contig46"  
 misc\_feature /note="assembly\_name:Contig47"  
 misc\_feature /note="assembly\_name:Contig48"  
 misc\_feature /note="assembly\_name:Contig49"  
 misc\_feature /note="assembly\_name:Contig50"  
 misc\_feature /note="assembly\_name:Contig51"  
 misc\_feature /note="assembly\_name:Contig52"  
 misc\_feature /note="assembly\_name:Contig53"  
 misc\_feature /note="assembly\_name:Contig54"  
 misc\_feature /note="assembly\_name:Contig55"

## Alignment Scores:

Pred. No.: 2.18e+03 Length: 173772  
 Score: 53.00 Matches: 11  
 Percent Similarity: 77.78% Conservatives: 3  
 Best Local Similarity: 61.11% Mismatches: 4  
 Query Match: 41.09% Indels: 0  
 DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC087862 (1-173772)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLaval 21  
 |||||  
 Db 173045 CTGGAGGCTCTGCTGCACAGACAGCGGCGACAGAGCAAGGAGTAGCCTTG 172992

## RESULT 19

LOCUS DME17920/c

DEFINITION Drosophila melanogaster mRNA for callosin protein.  
 ACCESSION Y17920  
 VERSION Y17920.1 GI:3893108

KEYWORDS calo gene.

SOURCE fruit fly.

## ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 12373)

AUTHORS Xu,X.Z., Wes,P.D., Chen,H., Li,H.S., Yu,M., Morgan,S., Liu,Y. and

Montell,C.

TITLE Retinal targets for calmodulin include proteins implicated in

synaptic transmission

J. Biol. Chem. 273 (47), 31297-31307 (1998)

MEDLINE 99030403

REFERENCE 2 (bases 1 to 12373)

AUTHORS Montell,C.

TITLE Direct Submission

Submitted (11-AUG-1998) C. Montell, The Johns Hopkins University

School of Medicine, Department of Biological Chemistry, 725 N.

Wolfe Street, Baltimore, MD 21205-2185, USA

## FEATURES

source

source

source

source

source

source

source

source

source

source

source

source

```

repeat_region 31133..31165
/rpt_family="ERVK"
repeat_region 31172..31597
/rpt_family="ERVK"
repeat_region 31598..32237
/rpt_family="ERV1"
repeat_region 34473..35285
/rpt_family="L1"
repeat_region 35883..36498
/rpt_family="ERV1"
repeat_region 38463..38929
/rpt_family="MERL_type"
repeat_region 39354..39609
/rpt_family="L1"
repeat_region 42594..42877
/rpt_family="Alu"
repeat_region 43519..43817
/rpt_family="Alu"
misc_feature 45998..46226
/note="similar to EST AA909108 (NID:g3048513) om32d11.s1"
repeat_region 47950..48034
/rpt_family="L1"
repeat_region 48869..48985
/rpt_family="Alu"

Alignment Scores:
Pred. No.: 1.45e+03 Length: 117309
Score: 53.00 Matches: 11
Percent Similarity: 77.78% Conservative: 3
Best Local Similarity: 61.11% Mismatches: 4
Query Match: 41.09% Indels: 0
DB: 9 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AC007274 (1-117309)

Qy 4 LeuGlulAluAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
Db 96844 CTGGAGGCTCTGCTGGACAGCAGACAGCGGCGACAGAGCAAGGAAGTAGCCTTG 96897

RESULT 18
AC087862/c 173772 bp DNA linear HTG 19-FEB-2001
LOCUS AC087862 Homo sapiens chromosome RPC1-11 clone RP11-724H13, WORKING DRAFT
DEFINITION AC087862
ACCESSION AC087862.3 GI:12963066
VERSION AC087862.3
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 173772)
Waterston,R.H.
Direct Submission
Submitted (02-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Feb 18, 2001 this sequence version replaced gi:12745114.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0724H13
----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads

```

```

Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 150537 bases at least Q40
Consensus quality: 158418 bases at least Q30
Consensus quality: 162327 bases at least Q20
Insert size: 191000; agarose-fp
Quality coverage: 2.88 in Q20 bases; agarose-fp
Quality coverage: 3.33 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1282: contig of 1282 bp in length
* 1283 1382: gap of unknown length
* 1383 2426: contig of 1044 bp in length
* 2427 2526: gap of unknown length
* 2527 3572: contig of 1046 bp in length
* 3573 3672: gap of unknown length
* 3673 6159: contig of 2487 bp in length
* 6160 6259: gap of unknown length
* 6260 8183: contig of 1924 bp in length
* 8184 8284: gap of unknown length
* 8284 11014: contig of 2731 bp in length
* 11015 11114: gap of unknown length
* 11115 12517: contig of 1403 bp in length
* 12518 12617: gap of unknown length
* 12618 14315: contig of 1698 bp in length
* 14316 14415: gap of unknown length
* 14416 17095: contig of 2680 bp in length
* 17096 17195: gap of unknown length
* 17196 19263: contig of 2068 bp in length
* 19264 19363: gap of unknown length
* 19364 22747: contig of 3383 bp in length
* 22747 22846: gap of unknown length
* 22847 24580: contig of 1734 bp in length
* 24581 24680: gap of unknown length
* 24681 27136: contig of 2456 bp in length
* 27137 27236: gap of unknown length
* 27237 30151: contig of 2915 bp in length
* 30152 30252: gap of unknown length
* 30252 33090: contig of 2839 bp in length
* 33091 33190: gap of unknown length
* 33191 36912: contig of 3722 bp in length
* 36913 37012: gap of unknown length
* 37013 40616: contig of 3604 bp in length
* 40617 40716: gap of unknown length
* 40717 44847: contig of 4131 bp in length
* 44848 44947: gap of unknown length
* 44948 50238: contig of 5291 bp in length
* 50239 50338: gap of unknown length
* 50339 53196: contig of 2858 bp in length
* 53197 53296: gap of unknown length
* 53297 57538: contig of 4242 bp in length
* 57539 57638: gap of unknown length
* 57639 62554: contig of 4916 bp in length
* 62555 66368: gap of unknown length
* 66369 66469: gap of unknown length
* 66470 71647: contig of 5179 bp in length
* 71648 71747: gap of unknown length
* 71748 74947: contig of 3200 bp in length
* 74948 75048: gap of unknown length
* 75049 79238: contig of 4191 bp in length
* 79239 79339: gap of unknown length
* 79340 84833: contig of 5495 bp in length
* 84834 84933: gap of unknown length
* 84934 89804: contig of 4871 bp in length

```



**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

Waterston,R.  
 Direct Submission  
 Submitted (14-OCT-2000) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Oct 14, 2000 this sequence version replaced gi:6624109.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: sapiens@waterston.wustl.edu  
 ----- Summary Statistics  
 -----  
 Center project name: HL\_NH0105L10  
 -----

**NOTICE:** This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

**SOURCE INFORMATION:**

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenoe,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

**NEIGHBORING SEQUENCE INFORMATION:**

The clone sequenced to the left is RP11-305H21; the clone sequenced to the right is RP11-109F19, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-105L10; actual end is at base position 24605 of RP11-109F19.

**FEATURES**

source

1..117309  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="Y"  
 /map="Y"  
 /clone="RP11-105L10"  
 /clone\_lib="RPCI-11"  
 733..1035

repeat\_region

/rpt\_family="Alu"

repeat\_region

/rpt\_family="MIR"

repeat\_region

/rpt\_family="Alu"

repeat\_region

/rpt\_family="MaLR"

repeat\_region

/rpt\_family="MIR"

repeat\_region

/rpt\_family="MIR"

repeat\_region

/rpt\_family="Alu"

repeat\_region

6920..7254

repeat\_region  
 /rpt\_family="L2"  
 7363..7578  
 /rpt\_family="ERV"  
 7586..8065  
 /rpt\_family="ERV1"  
 8066..8419  
 /rpt\_family="ERV"  
 8762..8877  
 /rpt\_family="ERV1"  
 8883..8945  
 /rpt\_family="Alu"  
 8961..9178  
 /rpt\_family="ERV1"  
 9185..10437  
 /rpt\_family="ERV1"  
 10455..10673  
 /rpt\_family="ERV1"  
 10676..10831  
 /rpt\_family="Alu"  
 10829..11215  
 /note="similar to EST AW070186 (NID:96025184) xa09a03.x1"  
 10829..11250  
 /note="similar to EST A1806840 (NID:95393406) wf15h08.x1"  
 11262..11805  
 /rpt\_family="ERV1"  
 12808..13247  
 /note="similar to EST A1908793 (NID:96499473)"  
 12809..12925  
 /note="similar to EST A1908824 (NID:96499504)"  
 12952..13235  
 /rpt\_family="ERV"  
 13237..13580  
 /rpt\_family="MaLR"  
 14241..14593  
 /rpt\_family="MaLR"  
 16242..16534  
 /rpt\_family="Alu"  
 17189..17450  
 /note="similar to EST D62720 (NID:9966494)"  
 18011..18146  
 /rpt\_family="L1"  
 18705..18738  
 /rpt\_family="L1"  
 18739..20933  
 /rpt\_family="L1"  
 20934..21246  
 /rpt\_family="L1"  
 21247..21557  
 /rpt\_family="Alu"  
 21740..21881  
 /rpt\_family="Alu"  
 21983..22295  
 /rpt\_family="ERV1"  
 23119..23239  
 /rpt\_family="MIR"  
 23639..23934  
 /rpt\_family="L2"  
 24194..24335  
 /rpt\_family="MaLR"  
 24452..24562  
 /rpt\_family="L1"  
 24961..25528  
 /rpt\_family="ERVK"  
 25530..27490  
 /rpt\_family="ERVK"  
 27485..27807  
 /rpt\_family="ERV1"  
 27808..27914  
 /rpt\_family="ERV1"  
 27922..28018  
 /rpt\_family="ERV1"  
 27997..29043  
 /rpt\_family="ERV1"





```
CDS
5899..7500
/function="putative membrane protein"
/notes="ORF6"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC93954.1"
/db_xref="GI:17974219"
/translation="MTLTPLSKORPGTTPSPRASDAPPTASATTAASPAAPA
PETATAGTGRWSLPGGLTQAQATAGAIUALALAGALIGLYLSEFERVATFAHORL
GFATLKGRLFTVGVVDGILVLIALDLMLAWLRPIGWVFPVLLTGTATVILNAASA
APSAWATLVYATAHAIPVLFIAIVEVGKTAYDVRVVPDPAHNGPGVPLRWFLA
PAPTFALRMKILWATVAEAEVLDELKRLVYRMLEERYGSKAPSDVRLPLTWER
YGLTIAEALALPQAEERQALREAEDEARLAEAEKOEKRAAAEAELARLKAAGAOAA
RTEVNSTAEVQAAAEVAAAEAAEAETAEVQSEAAAEAEADRAAEHRAAEARRR
AAADAEAEAEAEAEALTAEDAEARTAAEAESTEERARASQAQDAEAEEQAAA
EAHRAAEAEAEAEAEDELTPSRKVRKVKVARMILAEAGHPENLPLESVMTACSV
SQPTASEYSAQAOLLAEGYTPAR"
BASE COUNT 1125 a 2999 c 2583 g 979 t
ORIGIN

Alignment Scores:
Pred. No.: 86.7 Length: 7686
Score: 53.00 Matches: 13
Percent Similarity: 65.38% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 9
Query Match: 41.09% Indels: 0
DB: 1 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x SC0414671 (1-7686)
Qy 3 AlateuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 22
Db 6964 GCCAGCGGGCGGCGGACGAGCGGCGTCCAGTCGGAGCGCGCGCGGCC 7023
Qy 23 LysAlaAspAsnAlaAla 28
Db 7024 GAACCGCGCGCGCGGCC 7041

RESULT 16
SCSCP2/c
LOCUS
DEFINITION Streptomyces coelicolor plasmid SCP2.
ACCESSION AL645771
VERSION AL645771.1 GI:17026158
KEYWORDS bldA regulation; GntR-family regulator; IS1648; partitioning protein ParA; plasmid transfer; Tn5714; TraA; TraB; transposase; TraX.

SOURCE
ORGANISM Streptomyces coelicolor.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 31317)
Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. and Hopwood, D. A.
Hopwood, D. A. (Eds.);
PRACTICAL STREPTOMYCETES GENETICS: 1-1;
John Innes Foundation (2001)
ISBN 0-7084-0623-8.
2 (bases 1 to 31317)
Haug, I., Brolle, D., Weissenborn, A., Kieser, T., Wohlleben, W. and Altenbuchner, J.
Unpublished
3 (bases 1 to 31317)
Bentley, S. D., Parkhill, J., Barrell, B. G. and Rajandream, M. A.
Direct Submission
Submitted (14-NOV-2001) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Plasmid supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
1. .31317
source

/organism="Streptomyces coelicolor"
/db_xref="taxon:1902"
1..31317
/organism="Streptomyces coelicolor A3(2)"
/db_xref="A3(2)"
/db_xref="taxon:100226"
/clone="plasmid SCP2"
425..1402
/notes="Insertion element IS1648 bounded by 32bp imperfect inverted repeat. Copies of this IS element are also present in the Streptomyces coelicolor chromosome (EMBL:AL023861 and AL031035)."
492..917
/gene="SCP2.01"
492..917
/gene="SCP2.01"
/notes="Identical to Streptomyces coelicolor putative IS1648 transposase SC38.15 or SC6A9.36 TR:069927 (EMBL:AL023861) (141 aa) fasta scores: E(): 0, 100.0% id in 141 aa. Contains Pfam match to entry PF01511 Transposase_6, Transposase. Complete protein may be translated via a frameshift into the downstream CDS."
/codon_start=1
/transl_table=11
/product="putative IS1648 transposase"
/protein_id="CAD11995.1"
/db_xref="GI:17026159"
/translation="MGRGLTNRWSLLEPHPLPLGGRRGMRNDHRTVNGILFRVRT
GVPWRDILPERYGSWKTYERHRRMSADGTWDRILQSVQADNLAGRIDWSMVGVDSTS
CRAHOHAAGARKTRPRVYPKKQRPGTTAPTAPTRSDGPGAA"
594..839
/gene="SCP2.01"
/notes="Pfam match to entry PF01511 Transposase_6, transposase, score 12.30, E-value 3.6e-05"
914..1372
/gene="SCP2.02"
914..1372
/gene="SCP2.02"
/notes="Identical to Streptomyces coelicolor putative IS1648 transposase SC38.16 or SC6A9.37 TR:069928 (EMBL:AL023861) (152 aa) fasta scores: E(): 0, 99.3% id in 152 aa. Complete protein may be translated via a frameshift from the upstream CDS."
/codon_start=1
/transl_table=11
/product="putative IS1648 transposase"
/protein_id="CAD11996.1"
/db_xref="GI:17026160"
/translation="MTCKIHLAGCGCRPLALLTPQGWGDAPOLVEVLDIRVPRPL
GGRTPRPHVSGDKAYSSRRNRYLRRIIRITIPKQQRANRRRGEGGRPAGF
DRDHYRRNEVERTINRLKNPRAVATRYDKRAYVFGTGTAAALRLWLQ"
complement(3033..4010)
/gene="SCP2.04c"
complement(3033..4010)
/gene="SCP2.04c"
/notes="Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD11997.1"
/db_xref="GI:17026161"
/translation="MSRRSLALPSTRTEPDADALEAEPKLAARSAGVVASLTG
ADLSTPLTVAGLPPTDYVAETVAPLNDQERYLDVCEQALHGPKRKYVWAGKALEVI
NRGLRYRTHETFTADYVTEYMDKRAHAYRIEGRPADLYSPGIDNVEQARLEAPV
LKBYGPEVTYLYKRGKELRGDRVTAAELSEARAAALPPPKHARLPDQVRLTVAAA
EGRAPLAPAEKPVQAQAEHQAEQVDEGVSQDQVDEGAELATLCAEAQAQORQIV
DRVGCGTTLAALLYDPCGCDHLRELQYQARTAYRARDTSGEQVADDA"
complement(4007..4939)
/gene="para"
/notes="SCP2.05c"
complement(4007..4939)
/gene="para"
/notes="Similar to Caulobacter crescentus chromosome
```

```

mRNA      join(7165..7303,8920..9534,10965..>11147)
          /gene="CG7901"
          /product="CT42545"
          /db_xref="FLYBASE:FBan0007901"
          /db_xref="FLYBASE:FBgn0038572"
          <7165..>11147
          /gene="CG7901"
          /map="90E3-90E5"
          /db_xref="FLYBASE:FBan0007901"
          /db_xref="FLYBASE:FBgn0038572"
          join(9142..9534,10965..11147)
          /gene="CG7901"
          /note="CG7901 gene product [alt 1]"
          /codon_start=1
          /db_xref="FLYBASE:FBan0007901"
          /db_xref="FLYBASE:FBgn0038572"
          /protein_id="AAF55499.1"
          /db_xref="GI:7300340"
          /translation="MNEALDPTIKSTSAATPTAASETTTTAASSVWETTTTIAAA
          TAAAEKNETATNNNSNTSGSISSNNIVIPASATNGIKESNLSNTTTTAAAV
          AAATVEGVAITSTIVTGGTPTLSSLANLKDNTPPYDAPPTPISKVLNITGTP
          IVRKEKQTSARYNASKNCELTALIPNESE"
          join(9142..9534,10965..11147)
          /gene="CG7901"
          /note="CG7901 gene product [alt 2]"
          /codon_start=1
          /db_xref="FLYBASE:FBan0007901"
          /db_xref="FLYBASE:FBgn0038572"
          /protein_id="AAF55500.1"
          /db_xref="GI:7300341"
          /translation="MNEALDPTIKSTSAATPTAASETTTTAASSVWETTTTIAAA
          TAAAEKNETATNNNSNTSGSISSNNIVIPASATNGIKESNLSNTTTTAAAV
          AAATVEGVAITSTIVTGGTPTLSSLANLKDNTPPYDAPPTPISKVLNITGTP
          IVRKEKQTSARYNASKNCELTALIPNESE"
          join(12280..12385,12562..13626,13976..14480,14769..14965,
          15039..15373,15432..16135,16241..16677)
          /gene="CG7913"
          /product="CT23900"

Alignment Scores:
Pred. No.:      1.93e+03      Length:      224896
Score:          54.00         Matches:      12
Percent Similarity: 56.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match:     41.86%         Indels:      0
DB:              3             Gaps:       0

US-09-847-539a-6_COPY_59_86 (1-28) x AE003721 (1-224896)
Qy      5  GluAlaLeuAlaAspGlnThraspAlaLeuGlnSerGluGluAlaValValLysAla 24
      ||| -|||||::: |||||::: |||||::: |||||::: |||||::: |||||
Db 13577 GAGGAGGAGGAGTACGAGGAGGAGCGCGCGGAGGAGGAGGAGTCCGCGCGC 13518

Qy      25 AspAsnAlaAla 28
      ||| |||||
Db 13517 GATGCTGCGGCG 13506

RESULT 15
SC0414671 SC0414671 7686 bp DNA circular BCT 19-DEC-2001
LOCUS Streptomyces coelicolor plasmid 2 (SCP2)*.
DEFINITION AJ414671
ACCESSION AJ414671.1 GI:17974213
VERSION ORF1; ORF2; ORF3; ORF4; ORF5; ORF6.
KEYWORDS Streptomyces coelicolor.
SOURCE Streptomyces coelicolor
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycesaceae; Streptomyces.
1 (sites)
REFERENCE Wilkison,C.J., Hughes-Thomas,Z.A., Martin (nee Rowe),C.J.,
AUTHORS Bohm,I., Mironenko,T., Deacon,M., Wheatcroft,M., Wirtz,G.,
Staunton,J. and Leadlay,P.F.
TITLE Increasing the efficiency of heterologous promoters in

actinomycetes
J. Mol. Microbiol. Biotechnol. In press
2 (bases 1 to 7686)
REFERENCE Wilkison,C.J.
AUTHORS Direct Submission
TITLE Submitted (28-SEP-2001) Wilkison C.J., Department of Biochemistry,
JOURNAL University of Cambridge, 80 Tennis Court Rd., Cambridge, CB2 1GA,
UNITED KINGDOM
FEATURES
source      1..7686
            /organism="Streptomyces coelicolor"
            /plasmid="2 SCP2*"
            /db_xref="taxon:1902"
            1261..1713
            /function="unknown"
            /note="ORF1"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein"
            /protein_id="CAC93949.1"
            /db_xref="GI:17974214"
            /translation="MSARPTPHQFDGKGYTLOAGSDVPLYSLGLAAAEWATLEWLR
            EHGAAGYVVPVPEELGEDYGCASKDTCEKALNRLVKGVLVVKPGPRSGSYQLNPLRYW
            EGAGSTVYNACRRMAPPRVAPDDRAMTRSAKSKPTIPATRRRAAGEYR"
            1710..2129
            /function="unknown"
            /note="ORF2"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein"
            /protein_id="CAC93950.1"
            /db_xref="GI:17974215"
            /translation="MTTVPVEGFNPERDLTAPSLYSLNLSAAQHCTLAWEDHGGLFD
            VIPVETVAEDCGNSVTHEALARLEALNLVRSAGLYRINARYFTLHPLELREM
            ITAALTDPPTVDDRARARPKVSNTRRRRTIRVS"
            2140..2179
            complement(2884..3219)
            /function="unknown"
            /note="ORF3"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein"
            /protein_id="CAC93951.1"
            /db_xref="GI:17974216"
            /translation="MTNELDTASAAQETLERSAAELVRLQVVRELRAATHGLVYRALD
            GMEPPYLVAMEGRFVGVHRHGPAYLHSHWVLYPYGDODEPTAEAPDFNGPFGTAA
            AASLKPGQA"
            complement(2884..3219)
            complement(3216..4013)
            /function="putative transcriptional regulator"
            /note="ORF4"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein"
            /protein_id="CAC93952.1"
            /db_xref="GI:17974217"
            /translation="MATRRKRGPGGYEETAAHFRRLMDSGELSPGDPLPSMRDVCDOF
            GSATTVNRAFLQEEERTVSKPGVGTIVRDMRSRVVPFSTYGDVLAPGDKGQWER
            ATAAGDGRMLVEAPEVGGAPADVAARLGEPLGALVVRERRATIGEDVVLODADWY
            PLEATRAAGLDPRKPGVGGVLCAMTGAGLSTPTSDHDEVEVWVPSAQQAELSLSRVS
            VLVVERVYDTRVRLTHTGAADRLITLYKGLPLKATGAGSTS"
            4425..4904
            /function="unknown"
            /note="ORF5"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein"
            /protein_id="CAC93953.1"
            /db_xref="GI:17974218"
            /translation="MTRSVTRARRAFRTVADVRSTVAPASLVRSRADETAPDPTGL
            LAPEIEARAEIEADAADYAAATDRAESGRGKRKARKLDRLPAGRYGVMTVERVP
            SSRQTPDLEAIRATYKRLGLGDVPMTCAPSLRVTRAADVVPPEADEATAQLAA"
            4924..4947
            terminator
            CDS
            terminator
```

```

Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos-G.L., Abri,J.P., Agbayani,A., An,H.J., Basu,A.,
Andrews-PlannKoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beonkov,S., Borkova,D.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borikova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
Fleischmann,W., Fosler,C., Gabrielian,A.E., Garq,N.S.,
Gelbart,W.M., Glaesser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattel,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.Y., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Klanos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 224896)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
FEATURES             Location/Qualifiers
     1..224896
     /organism="Drosophila melanogaster"
     /db_xref="taxon:7227"
     /chromosome="3R"
     complement(join(1202..2465,2540..3150,3540..3806))
     /gene="CG7218"
     /product="Ct2261"
     /db_xref="FLYBASE:FBan0007218"
     /db_xref="FLYBASE:FBgn0038569"
     complement(<1202..>3806)
     /gene="CG7218"
     /map="90E2-90E4"
     /db_xref="FLYBASE:FBan0007218"
     /db_xref="FLYBASE:FBgn0038569"
     complement(join(1245..2465,2540..3150,3540..3738))
     /gene="CG7218"
     /note="CG7218 gene product"
     /codon_start=1
     /db_xref="FLYBASE:FBan0007218"
     /db_xref="FLYBASE:FBgn0038569"
     /protein_id="AAF53495.2"
     /db_xref="GI:10726596"
     /translation="MNATLNSAGKKRLRFRDGTGSRVEELHQQOEBQKAPLAQ
DIVAATPATPAQAQRLQSCATCTNTYDFKVTMTGCMLEHDEERYSAPRQ
KIYSPMRPRDLERFWVGIMQCADSEFLYIHFTLPVRFVMAVLSRTVARIPLRS
SGRLLSPAEICDLKGVIMVTIMLLVDNTRVHIKKSQSIKLKLYIFYNLEVG
RLLSAGQDITDALFTATEPKNSKREHFGLTVHLTIYFLHSLGIMFQATCLN
AVNSNKGILLTITANNFVELGSKVEKFKNNLQLTGSDVRERFHLISLVFVIO
TNKEFDMSITQCVMLPCFVAVLFTIILIDWKHAFITRFNLPESIVREYTTSLAYD
MTQTRQKHAFSDHSDILVARRMCFIFPPLAVLLIKAIYTAVSFENLAALLFLAYLFA

```

TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 source

mRNA

gene

CDS

Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Swirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

# TITLE JOURNAL

## COMMENT

Submitted (03-FEB-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Jan 31, 2000 this sequence version replaced gi:6587780. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdp@fruitfly.berkeley.edu](mailto:bdp@fruitfly.berkeley.edu). All contigs in this submission meet the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently consists of 38 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 667: contig of 667 bp in length  
668 747: gap of unknown length  
748 1288: contig of 541 bp in length  
1289 1368: gap of unknown length  
1369 7705: contig of 6337 bp in length  
7706 7785: gap of unknown length  
7786 13266: contig of 5481 bp in length  
13267 13346: gap of unknown length  
13347 134516: contig of 121170 bp in length  
134517 134596: gap of unknown length  
134597 134588: contig of 862 bp in length  
134598 135538: gap of unknown length  
135539 135539: contig of 226 bp in length  
135539 135765: gap of unknown length  
135765 135844: gap of unknown length  
135845 136080: contig of 236 bp in length  
136081 136160: gap of unknown length  
136161 136683: contig of 523 bp in length  
136684 137566: contig of 803 bp in length  
137567 137646: gap of unknown length  
137647 138152: contig of 506 bp in length  
138153 138232: gap of unknown length  
138233 138485: contig of 253 bp in length  
138486 138565: gap of unknown length  
138566 139286: contig of 721 bp in length  
139287 139366: gap of unknown length  
139367 140013: contig of 647 bp in length  
140014 140093: gap of unknown length  
140094 140815: contig of 722 bp in length  
140816 140895: gap of unknown length  
140896 141513: contig of 618 bp in length  
141514 141593: gap of unknown length  
141594 142372: contig of 779 bp in length  
142373 142452: gap of unknown length  
142453 143112: contig of 660 bp in length  
143113 143192: gap of unknown length  
143193 144042: contig of 850 bp in length  
144043 144122: gap of unknown length  
144123 144683: contig of 561 bp in length  
144684 147683: gap of unknown length  
147684 145169: contig of 406 bp in length  
145170 145249: gap of unknown length  
145250 146124: contig of 875 bp in length  
146125 146204: gap of unknown length  
146205 146205: contig of 775 bp in length  
146206 147059: gap of unknown length  
147060 147581: contig of 522 bp in length  
147582 147661: gap of unknown length  
147662 148202: contig of 541 bp in length  
148203 148282: gap of unknown length  
148283 148505: contig of 223 bp in length  
148506 148585: gap of unknown length

\* 148586 149025: contig of 440 bp in length  
149026 149105: gap of unknown length  
149106 149495: contig of 390 bp in length  
149496 149575: gap of unknown length  
149576 150161: contig of 586 bp in length  
150162 150241: gap of unknown length  
150242 150736: contig of 495 bp in length  
150737 150816: gap of unknown length  
150817 151159: contig of 343 bp in length  
151160 151239: gap of unknown length  
151240 152123: contig of 884 bp in length  
152124 152203: gap of unknown length  
152204 152586: contig of 383 bp in length  
152587 152666: gap of unknown length  
152667 153590: contig of 924 bp in length  
153591 153670: gap of unknown length  
153671 154319: contig of 649 bp in length  
154320 154399: gap of unknown length  
154400 154743: contig of 344 bp in length  
154744 154823: gap of unknown length  
154824 155022: contig of 599 bp in length  
155023 155502: gap of unknown length  
155503 155840: contig of 338 bp in length.

## FEATURES source

location/Qualifiers  
1..155840  
/organism="Drosophila melanogaster"  
/strain="y: cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="3"  
/map="90E1-90F10"  
/clone="BACR48113 (D522) RPCI-98 48.1.13"  
/clone\_lib="RPCI-98 (Roswell Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
pBACE3.6"

BASE COUNT 41735 a 34360 c 34802 g 41426 t 3517 others  
ORIGIN

Alignment Scores:  
Pred. No.: 1.32e+03 Length: 155840  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AC006495 (1-155840)

Oy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValysAla 24  
||| |||||: ||||| :|||: ||||| |||

Db 13996 GAGGAGGAGCTGACGAGGAGCGCCGCGAGGATGCGCGCGCTGCCGCC 13937  
||| |||||: ||||| :|||: ||||| |||

Oy 25 AspAsnAlaAla 28  
||| |||||

Db 13936 CATGCTGCCGCG 13925

RESULT 14  
AE003721/c

LOCUS AE003721 224896 bp DNA linear INV 05-OCT-2000  
DEFINITION Drosophila melanogaster genomic scaffold 142000013386035 section 46  
of 105, complete sequence.

ACCESSION AE003721 AE002708  
VERSION AE003721.1 GI:7300335

KEYWORDS HTG.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 224896)

ADDITIONAL INFORMATION  
Adams, M.D., Celnik, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,  
Ananides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galie, R.F.,  
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,  
Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,

For further information on this sequence e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

## FEATURES

source

Location/Qualifiers  
1..32270  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"

BASE COUNT 9248 a 7110 c 6761 g 9151 t  
ORIGIN

## Alignment Scores:

Pred. No.: 259 Length: 32270  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC014470 (1-32270)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallySala 24  
||| |||||: ||||| :|||:||||| |||

Db 26993 GAGGAGGAAGCTGACGAGGAGGACGCCGCGGAGGATGCCGCGGCTGCCGCC 26934  
||| |||||: ||||| :|||:||||| |||

QY 25 AspAsnAlaAla 28  
||| |||||

Db 26933 GATGCTGCCGCC 26922

## RESULT 12

AC009462/c

LOCUS

DEFINITION Drosophila melanogaster, chromosome 3R, region 90E-90E, BAC clone

BACK27G04, complete sequence.

AC009462

AC009462.6

GI:15451514

HTG.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 151610)

Celiker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,

Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,

Rogers,J.F., An,H., Baldwin,D., Banton,J., Beeson,K.Y., Bissam,D.A.,

Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,

Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,

Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,

Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,

Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,

McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,

Paclob,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,

Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,

Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,

Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

Sequencing of Drosophila chromosome 3R, region 90E-90E

Unpublished

2 (bases 1 to 151610)

Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,

Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,

Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,

Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,

Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,

Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclob,J.M., Park,S.,

Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,

Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and

Rubin,G.M.

Direct Submission

Submitted (23-AUG-1999) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Sep 6, 2001 this sequence version replaced gi:13384332.

COMMENT

Sequence submitted by:

Berkeley Drosophila Genome Project

Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome  
shotgun and from subclones of this BAC and its neighboring clones.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdgpe@fruitfly.berkeley.edu](mailto:bdgpe@fruitfly.berkeley.edu).

## FEATURES

source

1..151610

/organism="Drosophila melanogaster"

/strain="y: cn bw sp"

/db\_xref="taxon:7227"

/chromosome="3R"

/map="90E-90E"

/clone="BACR27G04 (D985)"

/clone\_lib="RPCI-98 (Roswell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI In

pbACe3.6)"

BASE COUNT 43139 a 33567 c 32376 g 42528 t

## ORIGIN

Alignment Scores:

Pred. No.: 1.29e+03 Length: 151610

Score: 54.00 Matches: 12

Percent Similarity: 66.67% Conservative: 4

Best Local Similarity: 50.00% Mismatches: 8

Query Match: 41.86% Indels: 0

DB: 3 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC009462 (1-151610)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallySala 24  
||| |||||: ||||| :|||:||||| |||

Db 73153 GAGGAGGAAGCTGACGAGGAGGACGCCGCGGAGGATGCCGCGGCTGCCGCC 73094  
||| |||||: ||||| :|||:||||| |||

QY 25 AspAsnAlaAla 28  
||| |||||

Db 73093 GATGCTGCCGCC 73082

## RESULT 13

AC006495/c

LOCUS

DEFINITION

Drosophila melanogaster chromosome 3 clone BACR48113 (D522) RPCI-98

48.1.13 map 90E1-90F10 strain y: cn bw sp, \*\*\* SEQUENCING IN

PROGRESS \*\*\*, 38 unordered pieces.

AC006495

AC006495.17 GI:6838793

HTG: HTGS\_PHASE1.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 155840)

Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,

Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,

Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,

Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,

Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,

Moshrefi,M., Nixon,K., Paclob,J.M., Park,S., Pfeiffer,B.,

Richards,S., Sethi,H., Svirskaas,R.R., Wan,K.H., Webster,D.,

Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.

Sequencing of Drosophila melanogaster

Unpublished

2 (bases 1 to 155840)

Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,

Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,

Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,

Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,

Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,

Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclob,J.M., Park,S.,

Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,

Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and

Rubin,G.M.

shows weak similarity to TR:AAF10061 (EMBL:AE001907), DR0482, Deinococcus radiodurans B-cell receptor associated protein-related protein (328 aa), fasta scores; E(): 4.9e-07, 23.4% identity in 265 aa overlap. Highly similar to TR:CA55231 (EMBL:AL117211), YPMT1.49C, Yersinia pestis CO-92 putative lipoprotein from plasmid pMT1 (276 aa), and TR:O68763 (EMBL:AF074611), Y1114, Yersinia pestis K1M5 scores; E(): 0, 97.5% identity in 280 aa overlap. Contains a possible N-terminal signal sequence and an appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site.  
/codon\_start=1  
/transl\_table=11  
/product="putative lipoprotein"  
/protein\_id="CAD09873.1"  
/db\_xref="GI:16505987"  
/translation="MRKGLLAVALTAICTMGLTGCDRVEFGYGVKVKLGDKGIGE VVGVGWGTGLNTELYFTFFKMTKYDEPFFOMSDGTAIGHKIGVAVLVNRKVTY PVQTYRGVDDITDLDLRQKIASLNLASRMRTDTFIDGKASLIDNALKDIOAEMS PVGTETVLSWVGKPDYPTQVIESINAKVTANKOTLQROQEVQRKAENMLREQAEG EADGEADAIRKRAQAEADAIKLRGEALRONPNVMELEAINKNWQLPQYMTGANTPE IALK"  
misc\_feature  
complement(2427..2459)  
/gene="HCN2.0006c"  
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"  
gene  
complement(2493..2693)  
/gene="HCN2.0007c"  
/note="HCN2.0007c"  
CDS  
complement(2493..2693)  
/gene="HCN2.0007c"  
highly similar to TR:CA55231 (EMBL:AL117211), YPMT1.49AC, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (66 aa), fasta scores; E(): 3.5e-25, 97.0% identity in 66 aa overlap. Contains hydrophobic, possible  
Alignment Scores:  
Pred. No.: 279 Length: 106516  
Score: 57.00 Matches: 13  
Percent Similarity: 68.18% Conservative: 2  
Best Local Similarity: 59.09% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 1 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x STYPPHC2 (1-106516)  
Qy 6 AlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAsp 25  
Db 39586 GCCGAGCATCACAAACAGCTGCATTACAAAGTGAGCAGGTTGCAGCGCCCGCATGCAGAC 39527  
Qy 26 AsnAla 27  
Db 39526 TCCGCG 39521  
RESULT 10  
AF031321  
LOCUS  
DEFINITION  
Tetrahymena thermophila granule lattice protein 5 precursor (GRL5)  
mRNA, complete cds.  
ACCESSION  
AF031321  
VERSION  
AF031321.1 GI:3366654  
KEYWORDS  
Tetrahymena thermophila.  
SOURCE  
Tetrahymena thermophila.  
ORGANISM  
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida; Tetrahymenina; Tetrahymena.  
REFERENCE  
1 (bases 1 to 1288)  
AUTHORS  
Verbisky,J.W. and Turkewitz,A.P.  
TITLE  
Proteolytic processing and Ca2+ binding activity of dense-core vesicle polypeptides in tetrahymena  
JOURNAL  
Mol. Biol. Cell (1997) In press

REFERENCE 2 (bases 1 to 1288)  
AUTHORS  
Verbisky,J.W. and Turkewitz,A.P.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (24-OCT-1997) Molecular Genetics and Cell Biology, University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA  
FEATURES  
Location/Qualifiers  
Source  
1..1288  
/organism="Tetrahymena thermophila"  
/strain="B1686"  
/db\_xref="taxon:5911"  
gene  
1..1288  
/gene="GRL5"  
/allele="wt"  
CDS  
24..1157  
/gene="GRL5"  
/note="Ca2+ binding protein; translated as a proprotein; Grl5p"  
/codon\_start=1  
/transl\_table=6  
/product="granule lattice protein 5 precursor"  
/protein\_id="AAC27989.1"  
/db\_xref="GI:3366655"  
/translation="MRVIAALVIALVCOSAMAVTSKQAKLMMEKINSKLEKSHLGR ALKGVITATKLYDYQDLYDAFAALKNOLLNLDNENSLFETQTASHDSAVAQFNAD ISNYGQINDAQAOLNDLSLNTYQONLQDAQAQADNDALNAEALANAEALYQ VATAEYSNADQVIGLAEKQEAQSHYDNADLSFSQVQIKNTFVFAQKVTATKGM NAKHQLFVRPVVQAMQVKNNTSSSIQTAIKALQDQAYFQKTFSDLTNEYVSFTQNM VQSTIDGLNQIIDIILQNOVTPGYEAQISSIQAOIOVEDALALAAQNLQDAQNALDAE NAGWENVVARHQALVDRINSEYDLVTQADRIVRNAQAQVNGSD"  
mat\_peptide  
195..1154  
/gene="GRL5"  
/note="Grl5p"  
/product="granule lattice protein 5"  
BASE COUNT 417 a 314 c 202 g 355 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 9.27 Length: 1288  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 3 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AF031321 (1-1288)  
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21  
Db 420 GATGCTCAACAGACTCTCTAAGACACACTGCTCTCAATGCTGCCGAGAGACTCTT 479  
Qy 22 VallysAlaAsp 25  
Db 480 GCCAATGCTGAA 491  
RESULT 11  
AC014470/c  
LOCUS  
DEFINITION  
Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
ACCESSION  
AC014470  
VERSION  
AC014470.1 GI:6436865  
KEYWORDS  
HTG; HTGS\_PHASE2.  
SOURCE  
fruit fly.  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 32270)  
AUTHORS  
Adams,M. and Venter,J.C.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
COMMENT  
This sequence was identified as CDM:10212984 by the submitter.

TITLE	Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whithead,S. and Barrell,B.G.	CDS	complement(375..650) /gene="HCM2.0003c"
JOURNAL	Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18		/note="HCM2.0003c, hypothetical protein, len: 91 aa; highly similar to TR:CAB62371 (EMBL:AL117211), YPMT1.468c, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (91 aa), fasta scores; E(): 0, 94.5% identity in 91 aa overlap and TR:Q92GX9 (EMBL:AF074611), Y1117, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (91 aa), fasta scores; E(): 0, 94.5% identity in 91 aa overlap"
PUBMED	Nature 413 (6858), 848-852 (2001)		/codon_start=1
REFERENCES	2 (bases 1 to 106516)		/transl_table=11
AUTHORS	Parkhill,J.		/product="hypothetical protein"
TITLE	Direct Submission		/protein_id="CAD09870.1"
JOURNAL	Submitted (25-OCT-2001) Submitted on behalf of the Salmonella sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk		/db_xref="GI:16505984"
COMMENT	Notes: Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web. (URL, <a href="http://www.sanger.ac.uk/Projects/S_typhi/">http://www.sanger.ac.uk/Projects/S_typhi/</a> ).	RBS	/translation="MKILVRISASTDYVYPLFMVKDCGLNDEETQAAIERNLVEYTGMDADSVYDDGVCVHNGSCWYVDDTMTVSDEDAHLERILGISTFE"
FEATURES	Location/Qualifiers		complement(658..663) /note="possible RBS"
source	1..106516	gene	complement(719..1129) /gene="HCM2.0004c"
	/organism="Salmonella enterica subsp. enterica serovar Typhi"	CDS	complement(719..1129) /gene="HCM2.0004c"
	/plasmid="pHCM2"		complement(719..1129) /gene="HCM2.0004c"
	/strain="Ct18"		complement(719..1129) /gene="HCM2.0004c"
	/db_xref="taxon:90370"		complement(719..1129) /gene="HCM2.0004c"
gene	complement(1..29)		complement(719..1129) /gene="HCM2.0004c"
CDS	/gene="HCM2.0001c"		complement(719..1129) /gene="HCM2.0004c"
	complement(1..29)		complement(719..1129) /gene="HCM2.0004c"
	/partial		complement(719..1129) /gene="HCM2.0004c"
	/gene="HCM2.0001c"		complement(719..1129) /gene="HCM2.0004c"
	/note="HCM2.0001c, hypothetical protein, len: 68 aa; highly similar to TR:CAB55227 (EMBL:AL117211), YPMT1.45c, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (71 aa), fasta scores; E(): 5.9e-25, 93.3% identity in 60 aa overlap and TR:O68760 (EMBL:AF074611), Y1119, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (71 aa), fasta scores; E(): 5.9e-25, 93.3% identity in 60 aa overlap. Spans the end of the sequence"		complement(719..1129) /gene="HCM2.0004c"
	/codon_start=1		complement(719..1129) /gene="HCM2.0004c"
	/transl_table=11		complement(719..1129) /gene="HCM2.0004c"
	/product="hypothetical protein"		complement(719..1129) /gene="HCM2.0004c"
	/protein_id="CAD09868.1"		complement(719..1129) /gene="HCM2.0004c"
	/db_xref="GI:16505982"		complement(719..1129) /gene="HCM2.0004c"
	/translation="MAVTLAGLE"		complement(719..1129) /gene="HCM2.0004c"
misc_feature	1..232		complement(719..1129) /gene="HCM2.0004c"
	/note="95.7% identical to Yersinia pestis CO-92 plasmid pMT1 (EMBL:AL117211) bases 48304..48335"		complement(719..1129) /gene="HCM2.0004c"
gene	complement(29..334)		complement(719..1129) /gene="HCM2.0004c"
CDS	/gene="HCM2.0002c"		complement(719..1129) /gene="HCM2.0004c"
	complement(29..334)		complement(719..1129) /gene="HCM2.0004c"
	/gene="HCM2.0002c"		complement(719..1129) /gene="HCM2.0004c"
	/note="HCM2.0002c, possible periplasmic protein, len: 101 aa; similar to TR:CAB55228 (EMBL:AL117211), YPMT1.46c, Yersinia pestis hypothetical protein from plasmid pMT1 (111 aa), fasta scores; E(): 2.8e-26, 95.5% identity in 67 aa overlap. The N-terminal 37 aa are highly similar to YPMT1.46ac (EMBL:AL117211) Yersinia pestis hypothetical protein from plasmid pMT1 (59 aa) (86.8% identity in 38 aa overlap) which lies immediately upstream of YPMT1.46c in pMT1. Contains a possible N-terminal signal sequence"		complement(719..1129) /gene="HCM2.0004c"
	/codon_start=1		complement(719..1129) /gene="HCM2.0004c"
	/transl_table=11		complement(719..1129) /gene="HCM2.0004c"
	/product="putative periplasmic protein"		complement(719..1129) /gene="HCM2.0004c"
	/protein_id="CAD09869.1"		complement(719..1129) /gene="HCM2.0004c"
	/db_xref="GI:16505983"		complement(719..1129) /gene="HCM2.0004c"
	/translation="MKTVLDTLLLLIVSTAFVLDCTFTCAIRKALAPVNGAMVVDGEGFYVVDRLVSSPYLTRKNGKLSGVGDKFYNKSGDYGVAHKNASYLFDCKEYV"		complement(719..1129) /gene="HCM2.0004c"
RBS	complement(36..43)		complement(719..1129) /gene="HCM2.0004c"
	/gene="HCM2.0002c"		complement(719..1129) /gene="HCM2.0004c"
	/note="possible RBS"		complement(719..1129) /gene="HCM2.0004c"
misc_feature	224..6722		complement(719..1129) /gene="HCM2.0004c"
	/note="96.5% identical to Yersinia pestis CO-92 plasmid pMT1 (EMBL:AL117211) bases 48733..55208"		complement(719..1129) /gene="HCM2.0004c"
gene	complement(375..650)		complement(719..1129) /gene="HCM2.0004c"
	/gene="HCM2.0003c"		complement(719..1129) /gene="HCM2.0004c"
	/note="HCM2.0003c, possible lipoprotein, len: 280 aa;		complement(719..1129) /gene="HCM2.0004c"





```
/gene="HP1328"
/Note="similar to GB:M26073 SP:P13510 PID:141927 percent
identity: 28.85; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="cation efflux system protein (czca)"
/protein_id="AAD08369.1"
/db_xref="GI:2314493"
/translation="MKRLWLLALILFSPFANAKTOEIKKTEAKSOTRFNISTTK
VIEKFSQSRYYALLPENEALIESQLRFGYVEKLYANKYTPPIKRGDRLLSYSP
ELVSAQELLSSLFNQVGAKEKLLGLGSSIEKLISSHQVNDMTIYHFNIGI
IFKPSQNEGSFKKQELFQIDLSQALWALVKNQDELEFKHAKIILFVEGKI
EQETTLLENIPINKKMLEARFNPNVKIYVPMFAQVEIFQPKMKILPKREAV
LKIGKATVPFKDDFGLSPLKIKAVRLSDGSYEILEGLKAGEEYVANNALFVLDADAQN
NGD"
7149..10256
/gene="HP1329"
CDS
7149..10256
/gene="HP1329"
/Note="similar to GB:M26073 SP:P13511 PID:141928 percent
identity: 31.32; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="cation efflux system protein (czca)"
/protein_id="AAD08370.1"
/db_xref="GI:2314494"
/translation="MIEKIIDLSVKNLLTTLVLLIFLASELWAIKSVRLDALPDLSP
AQQVQITYPNOSKIYQEQVYPLVSTFMSIANIDTVRGISSEGLIYIIFKDGVN
LYMARDRVLEQNLVSNLPKDAKVEIGSDSTSGWAOYALSSDKLSLKLQDDFY
RYVALLGVGVSEVASGVGFVDEVTLQNDLSLRYNLSLFOVANAINKSNNDTGGV
ILENGFEKIIRSHGYIQSLKDLLEIVYKKEAIPKIKINDIASVRLTPKPRGAANLNG
```

Alignment Scores:

Pred. No.:	30.4	Length:	12496
Score:	57.00	Matches:	12
Percent Similarity:	69.57%	Conservative:	4
Best Local Similarity:	52.17%	Mismatches:	7
Query Match:	44.19%	Indels:	0
DB:	1	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AE000634 (1-12496)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallys 23

Db 5041 ATACAGGCTTACAAGCAATTCAGCTTTAAGTTCTCAAGAAAAGTCGTTAGCAAG 5100

Qy 24 AlaAspAsn 26

Db 5101 TGGGATAAC 5109

RESULT 8

AE001548

LOCUS

DEFINITION Helicobacter pylori, strain J99 section 109 of 132 of the complete genome. BCT 20-JAN-1999

ACCESSION AE001548 AE001439

VERSION AE001548.1 GI:4155845

KEYWORDS

SOURCE Helicobacter pylori J99.

ORGANISM Helicobacter pylori J99

REFERENCE 1 (bases 1 to 12526)

AUTHORS Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guld, B.C., deJonghe, B.L., Carmel, G., Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F., and Trust, T.J.

TITLE Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori

JOURNAL Nature 397 (6715), 176-180 (1999)

99120557

MEDLINE

REMARK

Erratum: [[published erratum appears in Nature 1999 Feb 25:397(6721):719]]

2 (bases 1 to 12526)

AUTHORS King, B.L., Alm, R.A. and Trust, T.J.

TITLE Direct Submission

JOURNAL Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney Street, Cambridge, MA 02139, USA

COMMENT

Address all correspondence to: hp@arch.us.astro.com or Richard A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith, Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills, Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and Diane E. Taylor are with the University of Alberta Department of Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G 2H7 and the Canadian Bacterial Diseases Network. All other authors are with Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capability are available at ARCB's World Wide Web site. (URL: http://www.astro-boston.com/hpylori).

FEATURES

source

1..12526

/organism="Helicobacter pylori J99"

/strain="J99"

/db\_xref="taxon:85963"

114..1244

/gene="hpl1241"

114..1244

/gene="hpl1241"

/note="similar to H. pylori 26695 gene Hpl1321"

/codon\_start=1

/transl\_table=11

/product="putative"

/protein\_id="AAD06826.1"

/db\_xref="GI:4155853"

/translation="WRYDCLNPDPFIRKFMPLSCLHPFPETPKPALNNPLKKAHS

SDKICLLGPMKSGKTTFAKLAGKPNVYVNCNDRLNQNILSWLLKWLKMKMDL

LIDRIERLDFSLPKPLVILTPNYLTPNFSCVAGLGNFKEYISFPKPTPKN

TLFNRFLRDGNALDSFTENEQEKILKQENIQALQAFAPLMARKEYISQKFSVAFY

LYTOLKELKTSKDTLYLHLEKQRIIFLAPNFENHKTLYLCDFALPYSLTPSPS

LLNVFNMFVLELYKQPNVLYAHNCGIFILRNPTSKLALIAHAFTPTPHLEKQLLW

CHKHGFCLKIUVSINAPISANNPPYKHLNFIDFSLDIQSILV"

1509..2087

/gene="hpl1242"

1509..2087

/gene="hpl1242"

/note="similar to H. pylori 26695 gene Hpl1322"

/codon\_start=1

/transl\_table=11

/product="putative"

/protein\_id="AAD06827.1"

/db\_xref="GI:4155854"

/translation="MNALKKLSFCCALLSLGLFAOTVHAQHUKDTINYPDWLKLINLFDK

KNPNQYGSASISGKNDYFSNYIPYDDKLPPEKNAEIALRARNNAIYSLLESALL

TMCNRIYKALQVNNVYSHLDFGLVDFVTSKSLAKRFVDITNHRVYVWVQFPFIQPE

DLIAYFKAKRIDLSASATNLSAILNKALFHL"

complement(2061..2690)

/gene="rnhb"

/note="hpl1243"

complement(2061..2690)

/gene="rnhb"

/note="similar to H. pylori 26695 gene Hpl1323"

/codon\_start=1

/transl\_table=11

/product="RIBONUCLEASE HII"

/protein\_id="AAD06828.1"

/db\_xref="GI:4155855"

/translation="MGCVSMTLGLIDEAGRGCLAGSLFVAGVVCNEKIEFLKMKLKD

SKLSPKKRFEEDKIKTHGVGFVVKKSNIEDHGLGACLKLAIIEIVENGCSLA

NEIKIDGNTAFGLNKRYPNIQTIIKGDETIAQIAMSVAISKAKSDREMLIEIHALFKEY

GWKDKNGYGTQKHIEATNKLGCATPFHRHSFTLKNRILNPKLLEVEQRLV"

JOURNAL MEDLINE REMARK	Nature 388 (6642), 539-547 (1997) 97394467 Erratum:[published erratum appears in Nature 1997 Sep 25;389(6649):4121]]
REFERENCE	2 (bases 1 to 12496)
AUTHORS	Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G., Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S., Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F., Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G., Glodek,A., McKenney,K., Fitzgerald,L.M., Lee,N., Adams,M.D., Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R., Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C., Bowman,C., Watthey,L., Wallin,E., Hayes,W.S., Borodovsky,M., Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
TITLE	Direct Submission
JOURNAL	Submitted (06-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA
REFERENCE	3 (bases 1 to 12496)
AUTHORS	White,O.
TITLE	Direct Submission
JOURNAL	Submitted (17-MAR-1999) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1. .12496 /organism="Helicobacter pylori 26695" /strain="26695" /db_xref="taxon:85962" 113. .1246 /gene="Hpl321" 113. .1246 /gene="Hpl321" /note="similar to GB:L77117 PID:1592175 percent identity: 30.81; identified by sequence similarity; putative"
gene	/codon_start=1 /transl_table=11 /product="conserved hypothetical ATP-binding protein" /protein_id="AAD08366.1" /db_xref="GI:2314490"
CDS	/translation="MRYDYLNPISFERRKFMPLSCLHPFGFPETPKPVLNPLLRAPS SDKILGLPMGKGTTFALKLAKVFNPPVYINNDMLNONTILSSLLKWLKRLKMDL LILDRDLRDLFSIPLKPKIVLIPNCLSPATPNFSLCYALGLNKEYTSPFKPTPKN TLNRLRDLNADLSLETENEOKILKKQENIQIIFQAVAPLMARKICISQSKFVSAPY LYTLQRLKSLKDTLYKLKHALEKQRIPLFPNENNKTKLYLCDFALPYSUTSPS LLNVFNMFELYKOPPYELYSHDNGIFILRENSTNKIALIAHAAPTTPHFLKQLL WCHKHGLFNTIVVYSINAPISATNTPYKHLNFIQSLDLSIQSLV" 1452. .2057 /gene="Hpl322" 1452. .2057 /gene="Hpl322" /note="hypothetical protein; identified by Genemark; putative"
gene	/codon_start=1 /transl_table=11 /product="H. pylori predicted coding region Hpl322" /protein_id="AAD08374.1" /db_xref="GI:2314498"
CDS	/translation="MIFYKEATMNAKKLSFCALLSLGLFAQTAAHKLKGTINYPD WLEINFEENPPNQVGSASISGRNDYFANTIPYDQLPPEQNAEKIALLRARINA YSTLEILLKMHNRIVKVLQKNNVISHLFLGLVDLTSLKILAKRFVDVTNHRVYVM VQPFITQEDLIAYFAKAKRIDLSASATHLSALLNKALFHL" complement(2031. .2660) /gene="Hpl323" complement(2031. .2660) /gene="Hpl323" /note="similar to GB:L42023 SP:P43808 PID:1006316 PID:1221175 PID:1205308 percent identity: 36.32; identified by sequence similarity; putative"
gene	/codon_start=1 /transl_table=11 /product="ribonuclease HII (rnhB)" /protein_id="AAD08367.1" /db_xref="GI:2314491"
CDS	/translation="MGCVSMVTLGIDEAGRGCLAGSLFVAGVACNEKTALEFLKMLGD SKLSLKRRFFLEYIKTKHGEVGFVYVKKSANBIDSLGLGACLKLAQVBIENGCSLY DEIKDNTAFGLNKRYPHIQTIKGETIAQIAMASVLAKAFKQDREMLLEHALFREY GDKNGCYGTQKHIEAIKILGATPPHSHFTLKNRILNPKLLEVEQRLI" complement(2676. .2927) /gene="Hpl324" complement(2676. .2927) /gene="Hpl324" /note="hypothetical protein; identified by Genemark; putative"
gene	/codon_start=1 /transl_table=11 /product="H. pylori predicted coding region Hpl324" /protein_id="AAD08375.1" /db_xref="GI:2314499"
CDS	/translation="MNPETHCKEIKRKILKEALHDKNDLNLVLSGSKHAKLTSGAN HLIIPSSPSDRKSVKNFEKELTEFIKLRNRIHETCE" complement(2949. .4340) /gene="Hpl325" complement(2949. .4340) /gene="Hpl325" /note="similar to SP:P05042 GB:X00522 GB:X04065 PID:41513 GB:U00096 percent identity: 63.68; identified by sequence similarity; putative"
gene	/codon_start=1 /transl_table=11 /product="fumarase (fumC)" /protein_id="AAD08368.1" /db_xref="GI:2314492"
CDS	/translation="MQFRIEHDITMGEIKVNDQYMGACTORSLENEKIGTEKMPKELI GAFKLRSLAVNVNHLKGLSEKSAIKACDCILKGLCGEFPFIAIWOTGSGTQTN MNLNEVIANKATEILGNGFRKKLIHPNDVNMSSQSSNDTFTAMHIVSLEITHRL PSLENLKTFKESQOFKEIVKTRHLDQATPLTLGOEFSYASMLEHKSQOILFSL EHURELAIGTAVGTGLNHAIKELSEKVAELSQFSGVKFVAPNKFHALTSHDAIYA HGAFKALANLMIANDIWLASGPRGGLGELNIPENEGSSTMPCKVNPQCEAMTM VAYVGNDAIGIAASQGNFELNFKPVIYINFLQSLRLSSDSMSEFNHCASGLEP NREKIDYILHLSMLVLTALNPHVGENAAKIAKNAHKKGISRESALELKLJSAEDFD KFVYPEKMIGPKA" 4515. .4892 /gene="Hpl326" 4515. .4892 /gene="Hpl326" /note="hypothetical protein; identified by Genemark; putative"
gene	/codon_start=1 /transl_table=11 /product="H. pylori predicted coding region Hpl326" /protein_id="AAD08376.1" /db_xref="GI:2314500"
CDS	/translation="MRKLAALFLVSVGLVGNLNAWETLKAANDLEVKIKSVGNPIKGD NPIFILSPTLKGKALEKAIVRQFMPMPGMPAMKEMAQVSEKNGLYEAKTNLSMNGT WQVRVDTKSKEGOVYRAKTSIDL" 4897. .6135 /gene="Hpl327" 4897. .6135 /gene="Hpl327" /note="hypothetical protein; identified by Genemark; putative"
gene	/codon_start=1 /transl_table=11 /product="H. pylori predicted coding region Hpl327" /protein_id="AAD08377.1" /db_xref="GI:2314501"
CDS	/translation="MLSFISAFDKRGVSIRLLTALLLFLSLGAKDLQISQVAKYLS KNQIKALQEQIDALSOEKVSKWDNPILYLYGNANVSDFFRLQSTLMMNKSIGLS KQVDLNGKLTQSOMIDLEKQKILELTKTQOLATLSLMINGIENYKNOOEIELKTA IKNLENTLYQANHSSPNLJIAIAKLEILKSQLEIKKNNEALASASHYSNGELAFREN ELLISIAKPNFENRQELNLSATINDIAIARLDEEKSDQDITLAKKSPLEDNHTGCV YFPRQYNYDMFSEALSIPPIYQKAKLVEQKKESLVSFKSEVENTKNKTHLAL KLKLETLQKLNLSINKIIEKQEKIAQIYALDLKLSGNDYNAYNFAFNDRITITQITOL ETLSALNSTVLSLQNLKGLE" 6132. .7148 /gene="Hpl328" 6132. .7148

```

GTGTTSFVMTFLSNLVGVVYVYVNMVMDIKMFGSVMIFFIIMLFILARILILGAIF
NATVQEMSLGKLEGRSGDMIALKKTKLNGDSLSQSIEDSHTD"
complement(2980. .3840)
/ gene="map"
/ note="Spj1354"
complement(2980. .3840)
/ gene="map"
/ EC_number="3.4.11.18"
/ note="Best Blastp hit = spj0929J4|AMPM_BACHD METHIONINE
AMINOPEPTIDASE (MAP) (PEPTIDASE M)
>gi11264635|pir|IT44405 methionyl aminopeptidase [EC
3.4.11.18] map [imported] - Bacillus halodurans
>gi14512426|dbj|BAA75293.1| (AB017508) map homologue
(identity of 81% to B. subtilis ) [Bacillus halodurans]
>gi10172768|dbj|BAB03875.1| (AP001507) methionine
aminopeptidase [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative methionine aminopeptidase"
/protein_id="AAK34182.1"
/db_xref="GI:13622463"
/translacion="MILKSAEIEAMDRAGDFLAGIHLRLDIKPGVDMWEVAYV
RRKREDNVLPQIGVDGMMDDYPATCCGLNDEVAHAPPRHYILKGDLLKVDVYLS
EPLDKSIVDAALDFDNPENMKWTKGTSYGLGLADSCWAYAVGTPSDEIKQLMDVTKEA
MYRGTEKAVIGNRIGDICAAGVQEAESFGYGVRLVGHVGPTMHEDPMPVNYGTAG
RGLRLKGMVLTVEPMINTGWEIDTIKTGWAHKTLDGGLSCQYEHQVITKDGPIV
LTSQGEERTY"
complement(3856. .5139)
/ gene="Spj1355"
complement(3856. .5139)
/ gene="Spj1355"
/ note="Best Blastp hit = pir|A69998 hypothetical protein
ytoi - Bacillus subtilis >gi12293258|gb|AAC00336.1|
(AF008220) ytoi [Bacillus subtilis]
>gi12633392|emb|CAB14887.1| (Z99118) ytoi [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34183.1"
/db_xref="GI:13622464"
/translacion="MSKHQDILDYLEKLAIGKVSRSISNHLKVSDDGTAYRAIKEAE
NRGIVETPRSGTVRIKRGVRIDRLTYSEIARISDSEVLGAGHAGLEFSRFSIGA
MTQOINRRLYKVGGLIIVGDRETIQALLAHENHAILVGGFPVSRVTEMANNOIRPV
MYTHVDTTAVTMINHALSNIRIKTDLKTVEQVMLPITDYGLCEDSSVEERNTLIKK
TRQVRPVLDRKRVIGVVRDQVQDLPPTKLKVMKSNPITARPNTSLANISQMMI
FEDNLDPVDEENLGMITTRQAMENLPHQNPNTYSEQILSNLEEVVDYTOVV
VEPTMIDSAGNSMSVISEFLKEISIRALTKKHOKNIIIEQMVMVYFLHAIQIEDELKI
YPKITENRRSSTIDIEIFVDQVIKAKIITTKIN"
complement(5148. .5690)
/ gene="Spj1356"
complement(5148. .5690)
/ gene="Spj1356"
/ note="Best Blastp hit = emb|CAA03913.1| (AJ000084)
putative acetyl transferase [Proteus mirabilis]"
/codon_start=1
/transl_table=11
/product="putative acetyl transferase"
/protein_id="AAK34184.1"
/db_xref="GI:13622465"
/translacion="MDIWTKLAVFAFFETPKVILPFRFYEDHWDYFVSMVNDTKNLYYV
FPEOKTKAASDYLILVHSPFKPLGOWATEDKATHOVIGSIRIEHYDAKTRCADIGYFL
NYAFWGQIGIEMTVYKLVYLSFEPFGLTKLRIIHLNENKASQKVAKKAGFQKTCFKG
SDRNTHKICITKMTQLTWDR"
complement(5928. .6581)
/ gene="grab"
/ note="Spj1357"
complement(5928. .6581)
/ gene="grab"
/ note="Best Blastp hit = gb|AAD26340.1|AF124401.1
(AF124401) GRAB precursor [Streptococcus pyogenes]"
/codon_start=1
/transl_table=11

```

```

/product="protein GRAB (protein G-related alpha 2M-binding
protein)"
/protein_id="AAK34185.1"
/db_xref="GI:13622466"
/translacion="MGKEIKVKCFLLRRSAFGLVAVSASVLVGSVTSVAVDSPIEQPRII
PNGGTLTNLGNAPKLAIRNEERAIDELKQAIIDKQKATTAIEAASSDALADQOT
DALQSEEAAYVKADNAASDALEALADQOTDALQSEEAAYVQSDAAWEKAATPIAL
DVKTKDTRKPVVKKEERONVNTLPTTGESNPFPTAAALAINVSTGLVSVSSCKEN"
complement(6939. .8198)
/ gene="mur2"
/ note="mur2"
complement(6939. .8198)
/ gene="mur2"
/ note="Best Blastp hit = spj19670|MUA2_BACSU
UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE 2
(ENOLPYRUVATE TRANSFERASE) (UDP-N-ACETYLGLUCOSAMINE
ENOLPYRUVYL TRANSFERASE) (EPT) >gi12127262|pir|G32354
UDP-N-acetylglucosamine 1-carboxyvinyltransferase [EC
2.5.1.7] mur2 - Bacillus subtilis
>gi1853767|emb|CAA89875.1| (Z49782)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis] >gi12636235|emb|CAB15727.1| (Z99122)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis] >gi12636247|emb|CAB15738.1| (Z99123)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="putative UDP-N-acetylglucosamine
1-carboxyvinyltransferase"
/protein_id="AAK34186.1"
Alignment Scores:
Pred. No.: 1.85e-11 Length: 10029
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6_COPY_59_86 (1-28) x AE006573 (1-10029)
Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
|||||
Db 5308 TCAGATGCGCTTAGACATTAGCGGATCAACAGACGCTTTACATCAGAGAGCTGCG 6249
Qy 21 ValValLysAlaAspAsnAlaAla 28
|||||
Db 6248 GTTGTAAAAGCGGATAACGCTGCT 6225
RESULT 7
AE000634 12496 bp DNA linear BCT 06-APR-1999
LOCUS Helicobacter pylori 26695 section 112 of 134 of the complete
DEFINITION genome.
ACCESSION AE000634 AE000511
VERSION AE000634.1 GI:2314489
KEYWORDS Helicobacter pylori 26695.
SOURCE Helicobacter pylori 26695
ORGANISM Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
Helicobacter.
REFERENCE 1 (bases 1 to 12496)
AUTHORS Tomb, J.-F., White, O., Kervage, A.R., Clayton, R.A., Sutton, G.G.,
Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S.,
Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F.,
Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G.,
Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D.,
Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R.,
Peterson, J.D., Kelley, J.M., Karp, P.D., Smith, H.O., Fraser, C.M. and
Venter, J.C.
The complete genome sequence of the gastric pathogen Helicobacter
pylori
TITLE

```

```
source      1. .832
/organism="Streptococcus pyogenes"
/strain="ATCC700294"
/db_xref="ATCC:700294"
/db_xref="taxon:1314"
sig_peptide 1. .99
CDS         1. .654
/note="grab"
/codon_start=1
/transl_table=11
/product="GRAB precursor"
/protein_id="AAD26338.1"
/db_xref="GI:4589079"
/transl_table="MGKEIKVKCFLLRSAGFLVAVSASVILGVSTVSADVSPIEQPRII
PNGTLNLGNAPKIALNEERAIIDELKQAIEDKEATTAIEAASDAALEADOT
DALQSEAAVVKADNAASDALEALADOTDALQSEAEVVOGDNAASDAWEKAATPIAL
DVKTKTKPKVVKKEERQNVNLTPTTGESNPFTAAALAMVSTGVLVVSSCKEN"
mat_peptide 100. .651
/misc_feature 100. .273
/note="Region: alpha2-macroglobulin-binding A domain"
misc_feature 274. .441
/note="Region: repeat motif"
misc_feature 442. .576
/note="Region: cell-wall attachment"
misc_feature 577. .651
/note="Region: membrane-spanning"
BASE COUNT 304 a 143 c 169 g 216 t
ORIGIN
Alignment Scores:
Pred. No.:      1.41e-12      Length:      832
Score:          129.00      Matches:      28
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
Db:              1          Gaps:      0

US-09-847-539A-6_COPY_59_86 (1-28) x AF124399 (1-832)
Oy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20
Db 274 TCAGATGCCCTTAAAGCATTACCGGATCAACAGACGCTTTACAATCAGAGAAGCTGCG 333
Oy 21 ValValLysAlaAspAsnAlaAla 28
Db 334 GTTGTAAAGCGGATAACGCTGCT 357

RESULT 6
AE006573/c
LOCUS
DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section 102 of 167 of
the complete genome.
ACCESSION AE006573 AE004092
VERSION AE006573.1 GI:13622459
KEYWORDS
SOURCE
ORGANISM
Streptococcus pyogenes M1 GAS.
Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
1 (bases 1 to 10029)
Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,O., Zhu,H., Song,L., White,J.,
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
21192684
MEDLINE
PUBMED
REFERENCE
2 (bases 1 to 10029)
Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,O., Zhu,H., Song,L., White,J.,
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)

Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Direct Submission
Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 SL Young Blvd.,
Oklahoma City, OK 73104, USA
Location/Qualifiers
1. .10029
/organism="Streptococcus pyogenes M1 GAS"
/strain="Sp370"
/serotype="M1"
/db_xref="taxon:160490"
/complement(135. .626)
/gene="arok"
/note="Spv1351"
/complement(135. .626)
/gene="arok"
/EC_number="2.7.1.71"
/note="Best Blastp hit = pir|D82693 shikimate kinase
xfi335 [imported] - xylella fastidiosa (strain 9a5c)
>gi|9106332|gb|AAF84144.1|AE003966_5 (AE003966) shikimate
kinase [Xylella fastidiosa]"
/codon_start=1
/transl_table=11
/product="putative shikimate kinase"
/protein_id="AAK34179.1"
/db_xref="GI:13622460"
/transl_table="MTKVLGFMGVGKTTVSKHLSMHCKMDAIIIEAKIGMSIAAFIE
QHGEIAPRTIESOVKLDLIFANDNSIIIVTGGVVVLOENKOLLKHNHOLLVVASPE
TLYQRLKHDKKSQRLFLKYSKEAFYEQRMVVFEGLSDLVIRVDRHRTPEEVANII
EGY"
/complement(619. .1911)
/gene="aroA"
/note="Spv1352"
/complement(619. .1911)
/gene="aroA"
/EC_number="2.5.1.19"
/note="Best Blastp hit = sp143905|AROA_LACLA
3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE
(5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE) (EPSP
SYNTHASE) (EPSPS) >gi|1075724|pir||S52580
3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)
- Lactococcus lactis >gi|683583|emb|CAA55180.1| (X78413)
5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus
lactis]"
/codon_start=1
/transl_table=11
/product="putative 3-phosphoshikimate
1-carboxyvinyltransferase"
/protein_id="AAK34180.1"
/db_xref="GI:13622461"
/transl_table="MKRMKLRNAGPLOGTIOVPGDKSISHRAVILGAVAKGETRVKG
LLGEDVLSTIQAFNGLVRIEKKDOLVIEGOGFGLNAPCOTLNMGNSCTSMRLIA
GILAQPFVSKMIGDESLSKRPMDRIYPLQMGVEISGETDROFPPLQLQGNRLQIP
ITVLTPISSAOVKSAILLAQAQGTQVKEITRNTHTEEMIQOFGRLIVDGRKIT
LVGPQQLTAQETIPTYGDISAAFLVAGLIIPGSELKKNVAPRTGILLEVKEKM
AQIVYEDMKKEQVTSIRVYSNMKGTIISGGLIPRLIDELPIIATLTAQQTCTIK
DAQELRVKETRIQVTDILNSMGANIKATADGMIIKGTPLYVCANTSTYGDHRTGMW
TAIAALLVKQGVHLDKEEAIMTSYPTFFKDLERLCHD"
/complement(2013. .2978)
/gene="Spv1353"
/complement(2013. .2978)
/gene="Spv1353"
/note="Best Blastp hit = dbj|BAB07074.1| (AP001518)
unknown conserved protein [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34181.1"
/db_xref="GI:13622462"
/transl_table="MAEKKWFDKVLKSKWQYEPQVFMRHLSAEMDLISAIAVAYLIL
TAFPLIVTAANIIFYLNDIADLLRLKONLKPDIERPASAIVENIESKPSGSLGVA
TATGLWNTSRSLTSLOKAINKAYGASQHRDFHGLVGLLSTLILFLJALALIFSIF
SKAAIQVLDKRHYHLSDNITTTIFLLLIQPIITVLLIIFVGLMLLYFLLPNVKIKRILP
```

REFERENCE 2 (bases 1 to 717)  
 AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden

FEATURES  
 source 1..717  
 /organism="Streptococcus pyogenes"  
 /strain="KT19"  
 /db\_xref="taxon:1314"  
 mat\_peptide 1..>717  
 /product="GRAB"  
 misc\_feature 1..174  
 CDS  
 /note="Region: alpha2-macroglobulin-binding A domain"  
 <1..>717  
 /note="grab"  
 /codon\_start=1  
 /transl\_table=11  
 /product="GRAB precursor"  
 /protein\_id="AAD26341.1"  
 /db\_xref="GI:4589085"  
 /translation="VDSPIEQPRIIPNGTTLNLGNAPKELALRNEERAIDELKKQA  
 IDEKATTAIEAASSDALEADQDALQSEEAHVQSDNAASDALEALADQDALQSE  
 EAAVVKADNAASDTLEALADQDALQSEEAHVQSDNAASDTLEALADQDALQSE  
 AAVKADNAASDTLEALADQDALQSEEAHVQSDNAASDTLEALADQDALQSE  
 TRPVKKERQNVNLTPTTGE"  
 misc\_feature 175..594  
 /note="Region: repeat motif"  
 misc\_feature 595..717  
 /note="Region: cell-wall attachment"  
 BASE COUNT 257 a 140 c 163 g 157 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,21e-12 Length: 717  
 Score: 129.00 Matches: 28  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 1 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AF124402 (1-717)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
 |||||  
 Db 259 AGTGACGCTTAGAGCATTTGGCGATCAACACAGCGCTTTACAATCAGAAGAGCTGCG 318  
 |||||

Qy 21 ValVallyAlaAspAsnAlaAla 28  
 |||||

Db 319 GTTGTTAAAGCGGATACGCTGCT 342

RESULT 4  
 AF124403  
 LOCUS AF124403 804 bp DNA linear BCT 14-AUG-2000  
 DEFINITION Streptococcus pyogenes strain AP49 GRAB precursor, gene, partial cds.  
 ACCESSION AF124403  
 VERSION AF124403.1 GI:4589086  
 KEYWORDS  
 SOURCE Streptococcus pyogenes.  
 ORGANISM Streptococcus pyogenes  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.  
 1 (bases 1 to 804)  
 Rasmussen,M., Muller,H.P. and Bjorck,L.  
 TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
 J. Biol. Chem. 274 (22), 15336-15344 (1999)  
 JOURNAL MEDLINE 99269061  
 PUBMED 10336419  
 REFERENCE 2 (bases 1 to 804)  
 AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
 TITLE Direct Submission

JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden

FEATURES  
 source 1..804  
 /organism="Streptococcus pyogenes"  
 /strain="AP49"  
 /db\_xref="taxon:1314"  
 mat\_peptide 1..>804  
 /product="GRAB"  
 misc\_feature 1..174  
 CDS  
 /note="Region: alpha2-macroglobulin-binding A domain"  
 <1..>804  
 /note="grab"  
 /codon\_start=1  
 /transl\_table=11  
 /product="GRAB precursor"  
 /protein\_id="AAD26342.1"  
 /db\_xref="GI:4589087"  
 /translation="VDSPIEQPRIIPNGTTLNLGNAPKELALRNEERAIDELKKQA  
 IDEKATTAIEAASSDALEADQDALQSEEAHVQSDNAASDALEALADQDALQSE  
 EAAVVKADNAASDTLEALADQDALQSEEAHVQSDNAASDALEALADQDALQSE  
 ASVVKADNAASDTLEALADQDALQSEEAHVQSDNAASDTLEALADQDALQSE  
 VQSDNAASDAWEKAATPIALDVKTKDKTPVVKKEERQNVNLTPTTGE"  
 misc\_feature 175..678  
 /note="Region: repeat motif"  
 misc\_feature 679..804  
 /note="Region: cell-wall attachment"  
 BASE COUNT 284 a 161 c 185 g 174 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,36e-12 Length: 804  
 Score: 129.00 Matches: 28  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 1 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AF124403 (1-804)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
 |||||  
 Db 511 AGTGACGCTTAGAGCATTTGGCGATCAACACAGCGCTTTACAATCAGAAGAGCTGCG 570  
 |||||

Qy 21 ValVallyAlaAspAsnAlaAla 28  
 |||||

Db 571 GTTGTTAAAGCGGATACGCTGCT 594

RESULT 5  
 AF124399  
 LOCUS AF124399 832 bp DNA linear BCT 14-AUG-2000  
 DEFINITION Streptococcus pyogenes strain ATCC700294 GRAB precursor, gene, complete cds.  
 ACCESSION AF124399  
 VERSION AF124399.1 GI:4589078  
 KEYWORDS  
 SOURCE Streptococcus pyogenes.  
 ORGANISM Streptococcus pyogenes  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.  
 1 (bases 1 to 832)  
 Rasmussen,M., Muller,H.P. and Bjorck,L.  
 TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
 J. Biol. Chem. 274 (22), 15336-15344 (1999)  
 JOURNAL MEDLINE 99269061  
 PUBMED 10336419  
 REFERENCE 2 (bases 1 to 832)  
 AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden

FEATURES  
 Location/Qualifiers

```

ORGANISM      Streptococcus pyogenes
               Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
               Streptococcus.
REFERENCE     1 (bases 1 to 468)
AUTHORS      Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE        Protein GRAB of streptococcus pyogenes regulates proteolysis at the
               bacterial surface by binding alpha2-macroglobulin
JOURNAL      J. Biol. Chem. 274 (22), 15336-15344 (1999)
MEDLINE      99269061
PUBMED       10336419
REFERENCE     2 (bases 1 to 468)
AUTHORS      Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE        Direct Submission
JOURNAL      Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
               Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES     Location/Qualifiers
               source
               1..468
               /organism="Streptococcus pyogenes"
               /strain="KTL3"
               /db_xref="taxon:1314"
               mat_peptide
               1..>468
               /product="GRAB"
               misc_feature
               1..174
               /note="Region: alpha2-macroglobulin-binding A domain"
               CDS
               <1..>468
               /note="grab"
               /codon_start=1
               /transl_table=11
               /product="GRAB precursor"
               /protein_id="AAD26340.1"
               /db_xref="GI:4589083"
               /translation="VDSPIEQPRIIPNGTILNIGNAPEKALRNEERAIDELKKQA
               IEDKATTAIEAASDALEADQTDALQSEEAAYVKADNAASDALEALADQTDALOS
               EEAEVQSDNAASDAWEKAATPIALDVKTKDTPVKRKEQRQNTLPTTGE"
               misc_feature
               175..342
               /note="Region: repeat motif"
               misc_feature
               343..468
               /note="Region: cell-wall attachment"
               BASE COUNT 178 a 91 c 100 g 99 t
               ORIGIN
               Alignment Scores:
               Pred. No.: 7 76e-13 Length: 468
               Score: 129.00 Matches: 28
               Percent Similarity: 100.00% Conservative: 0
               Best Local Similarity: 100.00% Mismatches: 0
               Query Match: 100.00% Indels: 0
               DB: 1 Gaps: 0
               US-09-847-539a-6_COPY_59_86 (1-28) x AFI24401 (1-468)
               QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAa 20
               Db 175 TCAGATGCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAAATCAGAAGAGCTGCG 234
               QY 21 ValValIysAlaAspAsnAlaAa 28
               Db 235 GTTGTAAAGCGGATAACGCTGCT 258
               RESULT 2
               AFI24400
               LOCUS      AFI24400
               DEFINITION Streptococcus pyogenes strain AFI GRAB precursor, gene, partial
               cds.
               ACCESSION AFI24400
               VERSION   AFI24400.1 GI:4589080
               KEYWORDS
               SOURCE    Streptococcus pyogenes.
               ORGANISM  Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
               Streptococcus.
               REFERENCE 1 (bases 1 to 469)
               AUTHORS   Rasmussen,M., Muller,H.P. and Bjorck,L.
               TITLE     Protein GRAB of streptococcus pyogenes regulates proteolysis at the
               bacterial surface by binding alpha2-macroglobulin
               JOURNAL   J. Biol. Chem. 274 (22), 15336-15344 (1999)
               MEDLINE   99269061
               PUBMED    10336419

```

```

TITLE        Protein GRAB of streptococcus pyogenes regulates proteolysis at the
               bacterial surface by binding alpha2-macroglobulin
JOURNAL      J. Biol. Chem. 274 (22), 15336-15344 (1999)
MEDLINE      99269061
PUBMED       10336419
REFERENCE     2 (bases 1 to 469)
AUTHORS      Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE        Direct Submission
JOURNAL      Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
               Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES     Location/Qualifiers
               source
               1..469
               /organism="Streptococcus pyogenes"
               /strain="API"
               /db_xref="taxon:1314"
               mat_peptide
               1..>469
               /product="GRAB"
               misc_feature
               1..174
               /note="Region: alpha2-macroglobulin-binding A domain"
               CDS
               <1..>469
               /note="grab"
               /codon_start=1
               /transl_table=11
               /product="GRAB precursor"
               /protein_id="AAD26339.1"
               /db_xref="GI:4589081"
               /translation="VDSPIEQPRIIPNGTILNIGNAPEKALRNEERAIDELKKQA
               IEDKATTAIEAASDALEADQTDALQSEEAAYVKADNAASDALEALADQTDALOS
               EEAEVQSDNAASDAWEKAATPIALDVKTKDTPVKRKEQRQNTLPTTGE"
               misc_feature
               175..342
               /note="Region: repeat motif"
               misc_feature
               343..469
               /note="Region: cell-wall attachment"
               BASE COUNT 178 a 90 c 100 g 101 t
               ORIGIN
               Alignment Scores:
               Pred. No.: 7 77e-13 Length: 469
               Score: 129.00 Matches: 28
               Percent Similarity: 100.00% Conservative: 0
               Best Local Similarity: 100.00% Mismatches: 0
               Query Match: 100.00% Indels: 0
               DB: 1 Gaps: 0
               US-09-847-539a-6_COPY_59_86 (1-28) x AFI24400 (1-469)
               QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAa 20
               Db 175 TCAGATGCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAAATCAGAAGAGCTGCG 234
               QY 21 ValValIysAlaAspAsnAlaAa 28
               Db 235 GTTGTAAAGCGGATAACGCTGCT 258
               RESULT 3
               AFI24402
               LOCUS      AFI24402
               DEFINITION Streptococcus pyogenes strain KTL9 GRAB precursor, gene, partial
               cds.
               ACCESSION AFI24402
               VERSION   AFI24402.1 GI:4589084
               KEYWORDS
               SOURCE    Streptococcus pyogenes.
               ORGANISM  Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
               Streptococcus.
               REFERENCE 1 (bases 1 to 717)
               AUTHORS   Rasmussen,M., Muller,H.P. and Bjorck,L.
               TITLE     protein GRAB of streptococcus pyogenes regulates proteolysis at the
               bacterial surface by binding alpha2-macroglobulin
               JOURNAL   J. Biol. Chem. 274 (22), 15336-15344 (1999)
               MEDLINE   99269061
               PUBMED    10336419

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:12:06 ; Search time 273.262 Seconds  
(without alignments)  
2144.253 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALEALADTDALQSEEAAYVKADNAA 28

Scoring table:

BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlh  
-O=7cgn2\_1/USPTO\_spool/US09847539/runat\_10102002\_092548\_3423/app\_query.fasta\_1.526  
-DB=GenEmbl -GFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539.ecgn\_1\_1182.erunat\_10102002\_092548\_3423 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pi.\*
- 26: em\_lo.\*
- 27: em\_sts.\*
- 28: em\_un.\*

- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129	100.0	468	1	AF124401 Streptococ
2	129	100.0	469	1	AF124400 Streptococ
3	129	100.0	717	1	AF124402 Streptococ
4	129	100.0	804	1	AF124403 Streptococ
5	129	100.0	832	1	AF124399 Streptococ
6	129	100.0	10029	1	AE006573 Streptococ
7	57	44.2	12496	1	AE000634 Helicobac
8	57	44.2	12526	1	AE001548 Helicobac
9	57	44.2	106516	1	STYPPHCM2 Salmone
10	54	41.9	1288	3	AF031321 Tetrahyme
11	54	41.9	32270	2	AC014470 Drosophil
12	54	41.9	151610	3	AC009462 Drosophil
13	54	41.9	155840	2	AC006495 Drosophil
14	54	41.9	244896	3	AE003721 Drosophil
15	53	41.1	7686	1	SC0414671 Streptom
16	53	41.1	31317	1	SCSCP2 AL645771 Streptom
17	53	41.1	117309	9	AC007274 Homo sapi
18	53	41.1	173772	2	AC087862 Homo sapi
19	52	40.3	12373	3	DME17920 Drosophila
20	52	40.3	16135	3	AF096897 Drosophil
21	52	40.3	17595	3	AF096896 Drosophil
22	52	40.3	36993	2	AC018007 Drosophil
23	52	40.3	133155	2	AC011910 Drosophil
24	52	40.3	169427	3	AC007416 Drosophil
25	52	40.3	281993	3	AE003619 Drosophil
26	52	40.3	303100	1	RME603646 Rhizobium
27	51	39.5	2151	3	AY061146 Drosophil
28	51	39.5	3838	3	PPAS7 M10130 Plasmodium
29	51	39.5	12423	3	AE005155 Halobacte
30	51	39.5	14941	1	AF072878 Tetrahyme
31	51	39.5	20217	1	AE000047 Mycoplasma
32	51	39.5	30561	1	SC1F2 AL031350 Streptom
33	51	39.5	96288	3	AC005749 Drosophil
34	51	39.5	105823	2	AC019871 Drosophil
35	51	39.5	135509	8	AC037426 Oryza sat
36	51	39.5	175115	3	AC105899 Drosophil
37	51	39.5	183898	3	AC092244 Drosophil
38	51	39.5	200050	1	AL591973 Listeria
39	51	39.5	202050	1	AL646065 Ralstonia
40	51	39.5	231450	1	AL596163 Listeria
41	51	39.5	324484	3	AE003584 Drosophil
42	50	38.8	1298	1	ECODCDA M90069 E.coli deox
43	50	38.8	1409	1	ECU11035 U11035 Escherichia
44	50	38.8	10224	1	AE005432 Escherich
45	50	38.8	10780	3	CEMYO2 X08066 Caenorhabd

ALIGNMENTS

RESULT 1  
AF124401  
LOCUS AF124401 468 bp DNA linear BCT 14-AUG-2000  
DEFINITION Streptococcus pyogenes strain KTL3 GRAB precursor, gene, partial cds.  
ACCESSION AF124401  
VERSION AF124401.1 GI:4589082  
KEYWORDS Streptococcus pyogenes.  
SOURCE



**THIS PAGE BLANK (USPTO)**

```

; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 191:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 191:
US-08-961-083-191

Alignment Scores:
Pred. No.:          97.9          Length:          1033
Score:              44.00         Matches:          11
Percent Similarity: 65.00%        Conservative:     2
Best Local Similarity: 55.00%      Mismatches:      7
Query Match:        34.11%         Indels:           0
DB:                  4,            Gaps:             0

US-09-847-539A-6_COPY-58_86 (1-28) x US-08-961-083-191 (1-1033)

Qy      8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValValysLeAspAsnAla 27
      ||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      8 GCAGGGCAGACAGATGCTCGCAATTTGAAAGGCGCGCAGTTACCAAGGAGGAAAGCA 67
      ||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: October 13, 2002, 04:45:47
Job time : 99,2834 secs

```

;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 172754  
;; LENGTH: 862  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-172754

Alignment Scores:  
Pred. No.: 79.4 Length: 862  
Score: 44.00 Matches: 9  
Percent Similarity: 65.22% Conservative: 6  
Best Local Similarity: 39.13% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-027-632-172754 (1-862)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
Db 46 ACACAGACACTCCAGAGTCTCAGCAGTCTCAGTCTCCTCTGCGAGGAGAAAGCAGAGA 105  
Qy 21 ValVallys 23  
Db 106 GTTCTAAAG 114

RESULT 38

US-10-027-632-172755  
;; Sequence 172755, Application US/10027632

;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827,139

;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; CURRENT FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 172755  
;; LENGTH: 862  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-172755

Alignment Scores:  
Pred. No.: 79.4 Length: 862  
Score: 44.00 Matches: 9  
Percent Similarity: 65.22% Conservative: 6  
Best Local Similarity: 39.13% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-027-632-172755 (1-862)  
Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
Db 46 ACACAGACACTCCAGAGTCTCAGCAGTCTCAGTCTCCTCTGCGAGGAGAAAGCAGAGA 105  
Qy 21 ValVallys 23  
Db 106 GTTCTAAAG 114

RESULT 39

US-10-155-881-5807  
;; Sequence 5807, Application US/10153881

;; GENERAL INFORMATION:  
;; APPLICANT: Dotson, Stanton B.  
;; APPLICANT: Kovalic, David K.  
;; APPLICANT: Liu, Jingdong  
;; APPLICANT: Lutfiyya, Linda L.  
;; APPLICANT: McIninch, James  
;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
;; TRANSCRIPTION IN PLANTS  
;; FILE REFERENCE: 38-21(15300)J  
;; CURRENT APPLICATION NUMBER: US/10/155,881  
;; CURRENT FILING DATE: 2002-05-22  
;; NUMBER OF SEQ ID NOS: 37595  
;; SEQ ID NO 5807  
;; LENGTH: 940  
;; TYPE: DNA  
;; ORGANISM: Zea mays  
US-10-155-881-5807

Alignment Scores:

Pred. No.: 87.8 Length: 940  
Score: 44.00 Matches: 9  
Percent Similarity: 67.86% Conservative: 10  
Best Local Similarity: 32.14% Mismatches: 9  
Query Match: 34.11% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-155-881-5807 (1-940)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
Db 335 GCGGAGCGCTGGCGGGTGTCTCCAAAGTACGAGGCGTCCGCGCAGCACACGAG 394  
Qy 21 ValVallysAlaAspAlaAlaAla 28  
Db 395 ATGTGGCGCGCTGCTCCGCGTCG 418

RESULT 40

US-08-961-083-191

;; Sequence 191, Application US/08961083

;; GENERAL INFORMATION:  
;; APPLICANT: Choi et. al.  
;; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
;; NUMBER OF SEQUENCES: 452  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Human Genome Sciences, Inc.  
;; STREET: 9410 Key West Avenue  
;; CITY: Rockville  
;; STATE: Maryland  
;; COUNTRY: USA  
;; ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE: 30-Oct-1997  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-918-995-23409 (1-546)

QY 1 SerAspAlaLeuGlnThrAspAlaLeuGlnSerGluGluAlaAa 20  
DB 159 GCAAGAGCTCTGCCCTTGGCGGAGGAGCCATCTTTGCAAAACCAACCAAG 218  
QY 21 ValVallyAlaAspAsnAlaAa 28  
DB 219 GTGGCTATGGCGTTCCAGCAGCA 242

## RESULT 34

PCT-US02-27884-4854/c  
; Sequence 4854, Application PC/TUS0227884  
; GENERAL INFORMATION:  
; APPLICANT: The Dow Chemical Company  
; TITLE OF INVENTION: Nucleic Acid Compositions Conferring Altered Metabolic Character  
; FILE REFERENCE: DOM-07611  
; CURRENT APPLICATION NUMBER: PCT/US02/27884  
; CURRENT FILING DATE: 2002-08-30  
; NUMBER OF SEQ ID NOS: 7560  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4854  
; LENGTH: 656  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
PCT-US02-27884-4854

Alignment Scores:  
Pred. No.: 57.9 Length: 656  
Score: 44.00 Matches: 11  
Percent Similarity: 60.00% Conservative: 1  
Best Local Similarity: 55.00% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x PCT-US02-27884-4854 (1-656)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallyAla 24  
DB 120 GAGCGCGATGCCAGCCGCCCGCCGCGCAGCAAGATGCGGCGCGTGAATGCC 61

## RESULT 35

PCT-US02-27884-5555/c  
; Sequence 5555, Application PC/TUS0227884  
; GENERAL INFORMATION:  
; APPLICANT: The Dow Chemical Company  
; TITLE OF INVENTION: Nucleic Acid Compositions Conferring Altered Metabolic Character  
; FILE REFERENCE: DOM-07611  
; CURRENT APPLICATION NUMBER: PCT/US02/27884  
; CURRENT FILING DATE: 2002-08-30  
; NUMBER OF SEQ ID NOS: 7560  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5555  
; LENGTH: 656  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
PCT-US02-27884-5555

Alignment Scores:  
Pred. No.: 57.9 Length: 656  
Score: 44.00 Matches: 11  
Percent Similarity: 60.00% Conservative: 1  
Best Local Similarity: 55.00% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x PCT-US02-27884-5555 (1-656)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallyAla 24  
DB 120 GAGCGCGATGCCAGCCGCCCGCCGCGCAGCAAGATGCGGCGCGTGAATGCC 61

## RESULT 36

US-10-027-632-172753  
; Sequence 172753, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 172753  
; LENGTH: 862  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-172753

Alignment Scores:  
Pred. No.: 79.4 Length: 862  
Score: 44.00 Matches: 9  
Percent Similarity: 65.22% Conservative: 6  
Best Local Similarity: 39.13% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-027-632-172753 (1-862)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAa 20  
DB 46 ACACAGACACTCCAGAGTCTCAGACTCAGACTCATTCTCTGCAGGAGCAAGCAGAGA 105

QY 21 ValVallyys 23

DB 106 GTTCTAAG 114

## RESULT 37

US-10-027-632-172754  
; Sequence 172754, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218

```
; NUMBER OF SEQ ID NOS: 350
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 137
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (398)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-217-489-137

Alignment Scores:
Pred. No.: 35 Length: 425
Score: 44.00 Matches: 10
Percent Similarity: 60.87% Conservative: 4
Best Local Similarity: 43.48% Mismatches: 9
Query Match: 34.11% Indels: 0
DB: 6 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-10-217-489-137 (1-425)
Oy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValLysAla 24
Db 101 CAGGCATTAGAGCCCAACAGAGTCTTCCAGGCCCAAGAGCAGCAGTATGATGGAT 160
Oy 25 AspAsnAla 27
Db 161 CAGAGGCCA 169

RESULT 31
US-10-221-279-4262/c
; Sequence 4262, Application US/10221279
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-046
; CURRENT APPLICATION NUMBER: US/10/221,279
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 12360
; SOFTWARE: Custom
; SEQ ID NO 4262
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-221-279-4262

Alignment Scores:
Pred. No.: 41.2 Length: 489
Score: 44.00 Matches: 10
Percent Similarity: 61.34% Conservative: 6
Best Local Similarity: 38.46% Mismatches: 10
Query Match: 34.11% Indels: 0
DB: 6 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-10-221-279-4262 (1-489)
Oy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValVal 22
Db 114 GCGACAGAGCTCTGGTGGAGCCACTAGTCCTCCAGGACGAGAAAGCGAGATCCTG 55
Oy 23 LysAlaAspAsnAlaAla 28
Db 54 ACCGCTGACGGCGACGA 37

RESULT 32
US-10-027-632-13710/c
; Sequence 13710, Application US/10027632
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13710
; LENGTH: 511
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-13710

Alignment Scores:
Pred. No.: 43.3 Length: 511
Score: 44.00 Matches: 8
Percent Similarity: 81.25% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 34.11% Indels: 0
DB: 7 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-10-027-632-13710 (1-511)
Oy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 17
Db 132 GAGGCCATGGAAGCAACGCCTGATCAAACTGACTTAACTGCAGAC 85

RESULT 33
US-09-918-995-23409
; Sequence 23409, Application US/09918995
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 23409
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(546)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-23409

Alignment Scores:
Pred. No.: 46.8 Length: 546
Score: 44.00 Matches: 10
Percent Similarity: 60.71% Conservative: 7
Best Local Similarity: 35.71% Mismatches: 11
Query Match: 34.11% Indels: 0
DB: 5 Gaps: 0
```

Percent Similarity: 63.64% Conservative: 3  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 34.88% Indels: 0  
DB: 8 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-60-360-039-44171 (1-3435)

QY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22  
DB 1345 GCCAGGAGCTGCTGCCAGCAGGAAGCGGCTGCAGGAAGCGGAGCTGCACAAATC 1404

QY 23 LysAla 24  
DB 1405 GCCGCG 1410

RESULT 27  
US-09-975-254-11911  
; Sequence 11911, Application US/09975254  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: Heck, Gregory R.  
; APPLICANT: La Rosa, Thomas J.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15309)B  
; CURRENT APPLICATION NUMBER: US/09/975,254  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US/09/263,191  
; PRIOR FILING DATE: 1999-03-05  
; NUMBER OF SEQ ID NOS: 31255  
; SEQ ID NO 11911  
; LENGTH: 237  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 700905432H1  
US-09-975-254-11911

Alignment Scores:  
Pred. No.: 17.8 Length: 237  
Score: 44.00 Matches: 9  
Percent Similarity: 52.38% Conservative: 2  
Best Local Similarity: 42.86% Mismatches: 10  
Query Match: 34.11% Indels: 0  
DB: 5 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-975-254-11911 (1-237)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24  
DB 153 GAGACCTTCCTTCACAAACTCTCCTNNCTTGAGAGCAAGCCNTNGCCATACNNAGC 212

QY 25 Asp 25  
DB 213 GAC 215

RESULT 28  
US-09-918-995-37583/c  
; Sequence 37583, Application US/09918995  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 37583  
; LENGTH: 370  
; TYPE: DNA  
; ORGANISM: Homo sapiens

US-09-918-995-37583

Alignment Scores:  
Pred. No.: 29.8 Length: 370  
Score: 44.00 Matches: 10  
Percent Similarity: 60.87% Conservative: 4  
Best Local Similarity: 43.48% Mismatches: 9  
Query Match: 34.11% Indels: 0  
DB: 5 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-918-995-37583 (1-370)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24  
DB 129 CAGGCATTAGAGCCCAACACAGGATGCTTTCCAGGCCAAGACGACGACGTAATGATG 70

QY 25 AspAsnAla 27  
DB 69 CAGAAGGCA 61

RESULT 29  
US-09-918-995-37603/c  
; Sequence 37603, Application US/09918995  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 37603  
; LENGTH: 373  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-37603

Alignment Scores:  
Pred. No.: 30.1 Length: 373  
Score: 44.00 Matches: 10  
Percent Similarity: 60.87% Conservative: 4  
Best Local Similarity: 43.48% Mismatches: 9  
Query Match: 34.11% Indels: 0  
DB: 5 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-918-995-37603 (1-373)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24  
DB 129 CAGGCATTAGAGCCCAACACAGGATGCTTTCCAGGCCAAGACGACGACGTAATGATG 70

QY 25 AspAsnAla 27  
DB 69 CAGAAGGCA 61

RESULT 30  
US-10-217-489-137  
; Sequence 137, Application US/10217489  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE OF INVENTION: PM045CIN  
; CURRENT APPLICATION NUMBER: US/10/217,489  
; CURRENT FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 09/757,030  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04



Score: 45.00 Matches: 11  
Percent Similarity: 61.54% Conservative: 5  
Best Local Similarity: 42.31% Mismatches: 10  
Query Match: 34.88% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-027-632-164207 (1-825)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20

Db 180 GCAGATGCTTTGGAGATGCTGCTGCACAGACTGAAGCCTGTTGTCAAACAGGCTGTCA 121

QY 21 ValValLysAlaAspAsn 26

Db 120 CCTATCAGAGTTTAAAT 103

#### RESULT 20

US-10-027-632-164208/c

; Sequence 164208, Application US/10027632

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; PRIOR FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 164208

; LENGTH: 825

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-164208

#### Alignment Scores:

Pred. No.: 48.9 Length: 825

Score: 45.00 Matches: 11

Percent Similarity: 61.54% Conservative: 5

Best Local Similarity: 42.31% Mismatches: 10

Query Match: 34.88% Indels: 0

DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-027-632-164208 (1-825)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20

Db 180 GCAGATGCTTTGGAGATGCTGCTGCACAGACTGAAGCCTGTTGTCAAACAGGCTGTCA 121

QY 21 ValValLysAlaAspAsn 26

Db 120 CCTATCAGAGTTTAAAT 103

#### RESULT 21

US-10-219-999-22677

; Sequence 22677, Application US/10219999

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Edgerton, Michael D

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Stein, Joshua  
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-10(52726)C  
; CURRENT APPLICATION NUMBER: US/10/219,999  
; CURRENT FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: US 60/324,109  
; PRIOR FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: US 60/312,544  
; PRIOR FILING DATE: 2001-08-15  
; NUMBER OF SEQ ID NOS: 63520  
; SEQ ID NO 22677  
; LENGTH: 950  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (200)..(538)  
; OTHER INFORMATION:  
US-10-219-999-22677

#### Alignment Scores:

Pred. No.: 57.6 Length: 950

Score: 45.00 Matches: 10

Percent Similarity: 63.64% Conservative: 4

Best Local Similarity: 45.45% Mismatches: 8

Query Match: 34.88% Indels: 0

DB: 6 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-219-999-22677 (1-950)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20

Db 188 TCGTCGGCGGTGGAGACGCGCGGCGGTGGAGCGGTGGCGGTGGAGCGCGGTG 247

QY 21 ValVal 22

Db 248 GTGGTG 253

#### RESULT 22

US-60-360-039-27473/c

; Sequence 27473, Application US/60360039

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Chen, Xianfeng

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)A

; CURRENT APPLICATION NUMBER: US/60/360,039

; CURRENT FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 27473

; LENGTH: 1557

; TYPE: DNA

; ORGANISM: Neurospora crassa

US-60-360-039-27473

#### Alignment Scores:

Pred. No.: 102 Length: 1557

Score: 45.00 Matches: 9

Percent Similarity: 66.67% Conservative: 7

Best Local Similarity: 37.50% Mismatches: 8

Query Match: 34.88% Indels: 0

DB: 8 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-60-360-039-27473 (1-1557)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21

Db 188 TCGTCGGCGGTGGAGACGCGCGGCGGTGGAGCGGTGGCGGTGGAGCGCGGTG 247



; PRIOR APPLICATION NUMBER: US 60/218,006

Pred. No.:	.48.9	Length:	825
------------	-------	---------	-----

US-09-847-539A-6\_COPY\_59\_86 (1-28) x PCT-US02-25766-4301 (1-128978)

QY 8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValValValVal 24

Db 63278 GCTGAACAACTGACTCTTAGAACCACTGAGGCACAGGTCTTCAAAGCG 63328

## RESULT 13

US-10-161-453-6

; Sequence 6, Application US/10161453

; GENERAL INFORMATION:

; APPLICANT: EXELIXIS, INC.

; TITLE OF INVENTION: LINKS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE

; FILE REFERENCE: EX02-084C

; CURRENT APPLICATION NUMBER: US/10/161,453

; PRIOR FILING DATE: 2002-06-03

; PRIOR APPLICATION NUMBER: US 60/296,076

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: US 60/328,605

; PRIOR FILING DATE: 2001-10-10

; PRIOR APPLICATION NUMBER: US 60/357,253

; PRIOR FILING DATE: 2002-02-15

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 6

; LENGTH: 128978

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-161-453-6

## Alignment Scores:

Pred. No.:	7.16e+03	Length:	128978
Score:	47.00	Matches:	10
Percent Similarity:	76.47%	Conservative:	3
Best Local Similarity:	58.82%	Mismatches:	4
Query Match:	36.43%	Indels:	0
DB:	7	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-161-453-6 (1-128978)

QY 8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValValVal 24

Db 63278 GCTGAACAACTGACTCTTAGAACCACTGAGGCACAGGTCTTCAAAGCG 63328

## RESULT 14

US-10-155-881-3217/c

; Sequence 3217, Application US/10155881

; GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Lufliyya, Linda L.

; APPLICANT: McIninch, James

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: US/10/155,881

; PRIOR FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595

; SEQ ID NO 3217

; LENGTH: 705

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(705)

; OTHER INFORMATION: unsure at all n locations

US-10-155-881-3217

## Alignment Scores:

Pred. No.:	26.4	Length:	705
Score:	46.00	Matches:	10
Percent Similarity:	69.23%	Conservative:	8

Best Local Similarity: 38.46% Mismatches: 8  
Query Match: 35.66% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-155-881-3217 (1-705)

QY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22

Db 476 GCTGTCGTCGCTGTTGGCCGCGAAGCGAGGCAGTGTCTGAGGATGATCCCGCTGGCGTC 417

QY 23 LysAlaAspAsnAlaAla 28

Db 416 GAGGCTGACAACGGCGCC 399

## RESULT 15

US-10-219-999-6030/c

; Sequence 6030, Application US/10219999

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Edgerton, Michael D

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Stein, Joshua

; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT

; FILE REFERENCE: 38-10(52726)C

; CURRENT APPLICATION NUMBER: US/10/219,999

; PRIOR FILING DATE: 2002-08-15

; PRIOR APPLICATION NUMBER: US 60/324,109

; PRIOR FILING DATE: 2001-09-21

; PRIOR APPLICATION NUMBER: US 60/312,544

; PRIOR FILING DATE: 2001-08-15

; NUMBER OF SEQ ID NOS: 63520

; SEQ ID NO 6030

; LENGTH: 1512

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1410)

; OTHER INFORMATION:

US-10-219-999-6030

## Alignment Scores:

Pred. No.:	64	Length:	1512
Score:	46.00	Matches:	10
Percent Similarity:	69.23%	Conservative:	8
Best Local Similarity:	38.46%	Mismatches:	8
Query Match:	35.66%	Indels:	0
DB:	6	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-219-999-6030 (1-1512)

QY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22

Db 1314 GCTGTCGTCGCTGTTGGCCGCGAAGCGAGGCAGTGTCTGAGGATGATCCCGCTGGCGTC 1255

QY 23 LysAlaAspAsnAlaAla 28

Db 1254 GAGGCTGACAACGGCGCC 1237

## RESULT 16

US-60-360-039-44539

; Sequence 44539, Application US/60360039

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Chen, Xianfeng

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)A

QY 22 VallysAlaAspAsn 26  
:::  
Db 2235 CTGGCGAGCATAAC 2221

## RESULT 9

US-09-440-302B-636

; Sequence 636, Application US/09440302B

; GENERAL INFORMATION:

; APPLICANT: Chenchik, Alex

; APPLICANT: Lukashov, Matvey E.

; TITLE OF INVENTION: Human Neurobiology Array

; FILE REFERENCE: CLON-006CIP11

; CURRENT APPLICATION NUMBER: US/09/440.302B

; PRIOR FILING DATE: 1999-11-17

; PRIOR APPLICATION NUMBER: 09/053,375

; PRIOR FILING DATE: 1998-03-31

; NUMBER OF SEQ ID NOS: 1193

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 636

; LENGTH: 6965

; TYPE: DNA

; ORGANISM: Homo Sapiens

US-09-440-302B-636

## Alignment Scores:

Pred. No.:	243	Length:	6965
Score:	47.00	Matches:	11
Percent Similarity:	70.00%	Conservative:	3
Best Local Similarity:	55.00%	Mismatches:	6
Query Match:	36.43%	Indels:	0
DB:	5	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-440-302B-636 (1-6965)

QY 4 LeuGluAlaLeuAlaAspClnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23

Db 1460 TTGGAAGCCTTAGCATCAGCAGGATGCCCTTTGGCTGAAGAACTCCAACTCTTAAG 1519  
|||||:||||| ||| ||||||||| :||||||| |||:|||||

## RESULT 10

US-10-135-696-3/c

; Sequence 3, Application US/10135696

; GENERAL INFORMATION:

; APPLICANT: YAN, Chunhua et al.

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CL001178DIV

; CURRENT APPLICATION NUMBER: US/10/135,696

; CURRENT FILING DATE: 2002-05-01

; PRIOR APPLICATION NUMBER: 09/813,817

; PRIOR FILING DATE: 2001-03-22

; PRIOR APPLICATION NUMBER: 09/978,197

; PRIOR FILING DATE: 2001-10-17

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 59065

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-135-696-3

## Alignment Scores:

Pred. No.:	2,9e+03	Length:	59065
Score:	47.00	Matches:	10
Percent Similarity:	76.47%	Conservative:	3
Best Local Similarity:	58.82%	Mismatches:	4
Query Match:	36.43%	Indels:	0
DB:	7	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-135-696-3 (1-59065)

QY 8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24

|||||:|||||:|||||:|||||:||||| ||| |||||||

Db 21239 GCTGAACAACTGACTCCTTAGAACACAGTGAGGCACAGGTCTTCAAAGCG 21189

## RESULT 11

PCT-US01-17423-6

; Sequence 6, Application PC/TUS0117423

; GENERAL INFORMATION:

; APPLICANT: EXELIXIS, INC.

; TITLE OF INVENTION: LINKS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE

; FILE REFERENCE: EX02-084C-PC

; CURRENT APPLICATION NUMBER: PCT/US01/17423

; CURRENT FILING DATE: 2002-06-03

; PRIOR APPLICATION NUMBER: US 60/296,076

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: US 60/328,605

; PRIOR FILING DATE: 2001-10-10

; PRIOR APPLICATION NUMBER: US 60/357,253

; PRIOR FILING DATE: 2002-02-15

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 6

; LENGTH: 128978

; TYPE: DNA

; ORGANISM: Homo sapiens

PCT-US01-17423-6

## Alignment Scores:

Pred. No.:	7,16e+03	Length:	128978
Score:	47.00	Matches:	10
Percent Similarity:	76.47%	Conservative:	3
Best Local Similarity:	58.82%	Mismatches:	4
Query Match:	36.43%	Indels:	0
DB:	1	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x PCT-US01-17423-6 (1-128978)

QY 8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24

|||||:|||||:|||||:|||||:||||| ||| |||||||

Db 63278 GCTGAACAACTGACTCCTTAGAACACAGTGAGGCACAGGTCTTCAAAGCG 63328

## RESULT 12

PCT-US02-25766-4301

; Sequence 4301, Application PC/TUS0225766

; GENERAL INFORMATION:

; APPLICANT: GENE LOGIC, INC.

; APPLICANT: MUNGER, William E

; APPLICANT: FAULK, Ronald

; APPLICANT: SUN, Hongwei

; APPLICANT: SASAI, Hitoshi

; APPLICANT: WAGA, Iwao

; APPLICANT: YAMAMOTO, Jun

; TITLE OF INVENTION: Gene Expression Profiles in Glomerular Diseases

; FILE REFERENCE: 44921-5068-WO

; CURRENT APPLICATION NUMBER: PCT/US02/25766

; CURRENT FILING DATE: 2002-08-14

; PRIOR APPLICATION NUMBER: US 60/311,837

; PRIOR FILING DATE: 2001-08-14

; NUMBER OF SEQ ID NOS: 13946

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4301

; LENGTH: 128978

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Genbank Accession No. AC002073

PCT-US02-25766-4301

## Alignment Scores:

Pred. No.:	7,16e+03	Length:	128978
Score:	47.00	Matches:	10
Percent Similarity:	76.47%	Conservative:	3
Best Local Similarity:	58.82%	Mismatches:	4
Query Match:	36.43%	Indels:	0
DB:	1	Gaps:	0

```
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40312
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Bacillus thuringiensis
US-60-360-039-40312
Alignment Scores:
Pred. No.: 25 6 Length: 1449
Score: 48.00 Matches: 10
Percent Similarity: 65.22% Conservative: 5
Best Local Similarity: 43.48% Mismatches: 8
Query Match: 37.21% Indels: 0
DB: 8 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-360-039-40312 (1-1449)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
:: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1225 GAAGTACTCGTGGCTTGGTGGTCAATAGAGCTCTAGAACGAATGGAGCCAGCAGCA 1284

Qy 22 Vallysala 24
::: |||||
Db 1285 ATTAAGGCA 1293

RESULT 6
US-10-179-131-3446
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 3446
; LENGTH: 6093
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-179-131-3446
Alignment Scores:
Pred. No.: 135 Length: 6093
Score: 48.00 Matches: 13
Percent Similarity: 70.83% Conservative: 4
Best Local Similarity: 54.17% Mismatches: 6
Query Match: 37.21% Indels: 1
DB: 7 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-10-179-131-3446 (1-6093)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla-va 21
||| :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
Db 2125 GATATATCAAGGCTTAGCAGATCCAGTTGACTCTCTGTTATCAACCTCAGCTGCCAAT 2184

Qy 21 lVallysAla 24
: |||||
Db 2185 TGTGAAGCG 2194

RESULT 7
```

```
US-60-360-039-40814/C
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40814
; LENGTH: 783
; TYPE: DNA
; ORGANISM: Bacillus halodurans
US-60-360-039-40814
Alignment Scores:
Pred. No.: 19.4 Length: 783
Score: 47.00 Matches: 10
Percent Similarity: 63.64% Conservative: 4
Best Local Similarity: 45.45% Mismatches: 8
Query Match: 36.43% Indels: 0
DB: 8 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-360-039-40814 (1-783)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
||| ||||| ||| ::| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 691 CTGGCTCTTTGGCTCACCAGACGACCGAATGGAACGAGGAGCAATTCCTGTTTCA 632

Qy 24 AlaAsp 25
|||
Db 631 ACTGAT 626

RESULT 8
US-60-360-039-27525/c
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 27525
; LENGTH: 4885
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-60-360-039-27525
Alignment Scores:
Pred. No.: 161 Length: 4885
Score: 47.00 Matches: 10
Percent Similarity: 60.00% Conservative: 5
Best Local Similarity: 40.00% Mismatches: 10
Query Match: 36.43% Indels: 0
DB: 8 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-360-039-27525 (1-4885)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2295 GACGGCCTCGAACGAGTAGCGAGACTACATAGTACACACCGACCGACAGATCTT 2236
```



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:41:57 ; Search time 65.2834 Seconds  
(without alignments)  
1453.830 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALALADQTDALQSEAAVVKADNAA 28

Scoring table: BLOSUM62

Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2385415 seqs, 1694839300 residues

Total number of hits satisfying chosen parameters: 4770830

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DB=exlth

-Q=/cn2\_1/USPTO\_spool/US09847539/runat\_10102002\_092550\_3509/app\_query.fasta.1.526  
-DB=Pending\_Patents\_NA\_New -OFMT=fastap -SUFFIX=rnpn -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cd1  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539.@CN\_1.1.186.@runat\_10102002\_092550\_3509 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FCAPOP=6 -FCAPEXT=7  
-YGAPOP=60 -YCAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Pending Patents\_NA\_New:\*  
1: /cn2\_6/ptodata/1/pna/PCT\_NEW\_COMB.seq:\*  
2: /cn2\_6/ptodata/1/pna/US06\_NEW\_COMB.seq:\*  
3: /cn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq:\*  
4: /cn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq:\*  
5: /cn2\_6/ptodata/1/pna/US09\_NEW\_COMB.seq:\*  
6: /cn2\_6/ptodata/1/pna/US10\_NEW\_COMB.seq:\*  
7: /cn2\_6/ptodata/1/pna/US10\_NEW\_COMB.seq2:\*  
8: /cn2\_6/ptodata/1/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
C 1	50	38.8	1679	US-10-219-999-12394	Sequence 12394, A
C 2	50	38.8	1679	US-10-219-999-8111	Sequence 8111, Ap
C 3	50	38.8	1783	US-10-155-881-3337	Sequence 3337, Ap
C 4	50	38.8	3257	US-60-360-039-30630	Sequence 30630, A
C 5	48	37.2	1449	US-60-360-039-40312	Sequence 40312, A
C 6	48	37.2	6093	US-10-179-131-3446	Sequence 3446, Ap
C 7	47	36.4	783	US-60-360-039-40814	Sequence 40814, A
C 8	47	36.4	4885	US-60-360-039-27525	Sequence 27525, A
C 9	47	36.4	6965	US-09-440-302B-636	Sequence 636, App
C 10	47	36.4	59065	US-10-135-696-3	Sequence 3, Appli

11	47	36.4	128978	1	PCT-US01-17423-6	Sequence 6, Appli
12	47	36.4	128978	1	PCT-US02-25766-4301	Sequence 4301, Ap
13	47	36.4	128978	7	US-10-161-453-6	Sequence 6, Appli
C 14	46	35.7	705	7	US-10-155-881-3217	Sequence 3217, Ap
C 15	46	35.7	1512	6	US-10-219-999-6030	Sequence 6030, Ap
C 16	46	35.7	1608	8	US-60-360-039-44539	Sequence 44539, A
C 17	46	35.7	2199	8	US-60-360-039-42394	Sequence 42394, A
C 18	45	34.9	825	7	US-10-027-632-164206	Sequence 164206, A
C 19	45	34.9	825	7	US-10-027-632-164207	Sequence 164207, A
C 20	45	34.9	825	7	US-10-027-632-164208	Sequence 164208, A
C 21	45	34.9	950	6	US-10-219-999-22677	Sequence 22677, A
C 22	45	34.9	1557	8	US-60-360-039-27473	Sequence 27473, A
C 23	45	34.9	2243	7	US-10-027-632-262301	Sequence 262301, A
C 24	45	34.9	2688	8	US-60-360-039-28506	Sequence 28506, A
C 25	45	34.9	2845	8	US-60-360-039-31265	Sequence 31265, A
C 26	45	34.9	3435	8	US-60-360-039-44171	Sequence 44171, A
C 27	44	34.1	237	5	US-09-975-254-11911	Sequence 11911, A
C 28	44	34.1	370	5	US-09-918-995-37583	Sequence 37583, A
C 29	44	34.1	373	5	US-09-918-995-37603	Sequence 37603, A
C 30	44	34.1	425	6	US-10-217-489-137	Sequence 137, App
C 31	44	34.1	489	6	US-10-221-279-4262	Sequence 4262, Ap
C 32	44	34.1	511	7	US-10-027-632-13710	Sequence 13710, A
C 33	44	34.1	546	5	US-09-918-995-23409	Sequence 23409, A
C 34	44	34.1	656	1	PCT-US02-27884-4854	Sequence 4854, Ap
C 35	44	34.1	656	1	PCT-US02-27884-5555	Sequence 5555, Ap
C 36	44	34.1	862	7	US-10-027-632-172753	Sequence 172753, A
C 37	44	34.1	862	7	US-10-027-632-172754	Sequence 172754, A
C 38	44	34.1	862	7	US-10-027-632-172755	Sequence 172755, A
C 39	44	34.1	940	7	US-10-155-881-5807	Sequence 5807, Ap
C 40	44	34.1	1033	4	US-08-961-083-191	Sequence 191, App
C 41	44	34.1	1216	6	US-10-217-550-124	Sequence 124, App
C 42	44	34.1	1428	7	US-10-155-881-17609	Sequence 17609, A
C 43	44	34.1	1484	6	US-10-219-999-26942	Sequence 26942, A
C 44	44	34.1	1811	6	US-10-219-999-30708	Sequence 30708, A
C 45	44	34.1	1890	8	US-60-360-039-44420	Sequence 44420, A

#### ALIGNMENTS

RESULT 1  
US-10-219-999-12394/c  
; Sequence 12394, Application US/10219999  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Edgerton, Michael D  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Stein, Joshua  
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-10(52726)C  
; CURRENT APPLICATION NUMBER: US/10/219,999  
; CURRENT FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: US 60/324,109  
; PRIOR FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: US 60/312,544  
; PRIOR FILING DATE: 2001-08-15  
; NUMBER OF SEQ ID NOS: 63520  
; SEQ ID NO 12394  
; LENGTH: 1679  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2)..(1351)  
; OTHER INFORMATION:  
US-10-219-999-12394  
Alignment Scores: 12.8 Length: 1679  
Pred. No.: 50.00 Matches: 11  
Score: 73.08% Conservative: 8  
Percent Similarity: 42.31% Mismatches: 7  
Best Local Similarity: 42.31%

**THIS PAGE BLANK (USPTO)**

```
; NAME/KEY: misc_feature
; LOCATION: (1)..(1230)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(1230)
; OTHER INFORMATION: 40182 : FINISHED (Clone Number : FINISHED)
US-09-565-309A-62025
```

```
Alignment Scores:
Pred. No.: 177 Length: 1230
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-565-309A-62025 (1-1230)

```
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23
||| |||||:||||: ||| ||| :|||:|||||: |||
Db 518 CTTCCGGCACTCTCCGACAAAGGACATGCTCAAAATTCAAAAGCTGCAATCGCAAG 459

Qy 24 AlaAsp 25
|||||
Db 458 GCGGAT 453
```

```
RESULT 40
US-09-595-329A-545/c
; Sequence 545, Application US/09595329A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
; FILE REFERENCE: 2750-0948P
; CURRENT APPLICATION NUMBER: US/09/595,329A
; CURRENT FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 3309
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 545
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Arabidopsis Thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1230)
; OTHER INFORMATION: Ceres Seq. ID no. 1013068
; NAME/KEY: misc_feature
; LOCATION: (1)..(1)
; OTHER INFORMATION: n is a, c, t, g, unknown, or other
US-09-595-329A-545
```

```
Alignment Scores:
Pred. No.: 177 Length: 1230
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-595-329A-545 (1-1230)

```
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23
||| |||||:||||: ||| ||| :|||:|||||: |||
Db 518 CTTCCGGCACTCTCCGACAAAGGACATGCTCAAAATTCAAAAGCTGCAATCGCAAG 459

Qy 24 AlaAsp 25
|||||
Db 458 GCGGAT 453
```

Search completed: October 13, 2002, 07:14:46  
Job time : 567.553 secs



```
RESULT 36
US-09-565-309A-67148/c
; Sequence 67148, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 67148
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(939)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(939)
; OTHER INFORMATION: 40182 : CDS (Clone Number : CDS)
US-09-565-309A-67148

Alignment Scores:
Pred. No.: 130 Length: 939
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-67148 (1-939)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLys 23
DB 382 CTTCGCGCACTCCGACAAAGGACATGCCTCAAAATTCAAAAAGCTGCAATCGCAAG 323

QY 24 AlaAsp 25
DB 322 GCGGAT 317

RESULT 37
US-09-708-427-29368/c
; Sequence 29368, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29368
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..984
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: 1..984
; OTHER INFORMATION: Ceres Seq. ID 1825505
US-09-708-427-29368

Alignment Scores:
Pred. No.: 137 Length: 984
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
```

```
Query Match: 40.31% Indels: 0
DB: 28 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-708-427-29368 (1-984)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLys 23
DB 442 CTTCGCGCACTCCGACAAAGGACATGCCTCAAAATTCAAAAAGCTGCAATCGCAAG 383

QY 24 AlaAsp 25
DB 382 GCGGAT 377

RESULT 38
US-09-565-309A-67147/c
; Sequence 67147, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 67147
; LENGTH: 999
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(999)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(999)
; OTHER INFORMATION: 40182 : CDS (Clone Number : CDS)
US-09-565-309A-67147

Alignment Scores:
Pred. No.: 139 Length: 999
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-67147 (1-999)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLys 23
DB 442 CTTCGCGCACTCCGACAAAGGACATGCCTCAAAATTCAAAAAGCTGCAATCGCAAG 383

QY 24 AlaAsp 25
DB 382 GCGGAT 377

RESULT 39
US-09-565-309A-62025/c
; Sequence 62025, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 62025
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
```

```
; APPLICANT: RUSSO, FRANK D.
; APPLICANT: HANN, AMY L.
; APPLICANT: HEATH, JOE D.
; APPLICANT: FINNEY, GREGORY L.
; APPLICANT: BROOKS, JACQUELINE
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF PSEUDOMONAS AERUGINOSA
; TITLE OF INVENTION: GENOME, FRAGMENTS THEREOF, AND USES THEREOF
; NUMBER OF SEQUENCES: 860
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/082,302
; FILING DATE: HEREWITH
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PM-0011-1 P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 513:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8830 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; IMMEDIATE SOURCE:
; CLONE: PAE1G517
; US-60-082-302-513

Alignment Scores:
Pred. No.: 1,41e+03 Length: 8830
Score: 52.50 Matches: 14
Percent Similarity: 71.43% Conservative: 6
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 40.70% Indels: 5
DB: 47 Caps: 1

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-082-302-513 (1-8830)

Qy 5 GluAlaLeuAlaAsp-----GlnThrAspAlaLeuGlnSerGluGluAla 19
|||||:|||||:
|:|||||:|||||:|||||:
Db 2929 GAAGCGGTAGTGTGAGTGAGCGAGCGGACTGCTGCGAGCGGATGTCGCC 2870

Qy 20 AlaValVallysAlaAspAsnAla 27
|||||:|||||:|||||:
Db 2869 GCTGTCGTAGCGCGGACCGCTGCC 2846

RESULT 34
US-09-565-309A-32821/C
; Sequence 32821, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 32821

Alignment Scores:
Pred. No.: 83.4 Length: 641
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-32824 (1-641)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVallys 23
||| |||||:|||||: ||| ||| :|||||: |||
Db 461 CTTCGGCAGCTCTCGGACAAAGGACATGCTCAAAATTCAAAAGCTGCAATGCCAAAG 402

Qy 24 AlaAsp 25
|||||
Db 401 GCGGAT 396
```

```
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(594)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(594)
; OTHER INFORMATION: 40182:107311 (Clone Number:Unique Sequence Identifier)
US-09-565-309A-32821

Alignment Scores:
Pred. No.: 76.4 Length: 594
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-32821 (1-594)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVallys 23
||| |||||:|||||: ||| ||| :|||||: |||
Db 518 CTTCGGCAGCTCTCGGACAAAGGACATGCTCAAAATTCAAAAGCTGCAATGCCAAAG 459

Qy 24 AlaAsp 25
|||||
Db 458 GCGGAT 453

RESULT 35
US-09-565-309A-32824/C
; Sequence 32824, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 32824
; LENGTH: 641
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(641)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(641)
; OTHER INFORMATION: 40182:957281 (Clone Number:Unique Sequence Identifier)
US-09-565-309A-32824

Alignment Scores:
Pred. No.: 83.4 Length: 641
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-32824 (1-641)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVallys 23
||| |||||:|||||: ||| ||| :|||||: |||
Db 461 CTTCGGCAGCTCTCGGACAAAGGACATGCTCAAAATTCAAAAGCTGCAATGCCAAAG 402

Qy 24 AlaAsp 25
|||||
Db 401 GCGGAT 396
```

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17589  
; LENGTH: 7206  
; TYPE: DNA  
; ORGANISM: Drosophila  
US-60-167-217-17589

Alignment Scores:  
Pred. No.: 619 Length: 7206  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 55 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-60-167-217-17589 (1-7206)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAla 24  
DB 4905 GAGGAGGAAGCTGACGAGGAGGACGCCGCGGAGGAGGATGCCGCGGCTGCCGCC 4964  
QY 25 AspAsnAlaAla 28  
DB 4965 GATGCTGCGCGC 4976

RESULT 30

US-09-528-237A-1811/c  
; Sequence 1811, Application US/09528237A

; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; TITLE OF INVENTION: The Drosophila Genome, Primary Nucleic  
; TITLE OF INVENTION: Acid Sequences, Systems Containing The Nucleic Acid  
; FILE OF INVENTION: Sequences and Uses Thereof  
; FILE REFERENCE: CL000284  
; CURRENT APPLICATION NUMBER: US/09/528,237A  
; CURRENT FILING DATE: 2000-03-17  
; NUMBER OF SEQ ID NOS: 2925

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1811  
; LENGTH: 32275  
; TYPE: DNA  
; ORGANISM: Drosophila  
US-09-528-237A-1811

Alignment Scores:  
Pred. No.: 3,48e+03 Length: 32275  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 19 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-528-237A-1811 (1-32275)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAla 24  
DB 26995 GAGGAGGAAGCTGACGAGGAGGACGCCGCGGAGGAGGATGCCGCGGCTGCCGCC 26936  
QY 25 AspAsnAlaAla 28  
DB 26935 GATGCTGCGCGC 26924

RESULT 31

US-09-252-991A-8951/c  
; Sequence 8951, Application US/09252991A

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8951

; LENGTH: 1239  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8951

Alignment Scores:  
Pred. No.: 147 Length: 1239  
Score: 52.50 Matches: 14  
Percent Similarity: 71.43% Conservative: 6  
Best Local Similarity: 50.00% Mismatches: 3  
Query Match: 40.70% Indels: 5  
DB: 16 Gaps: 1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-252-991A-8951 (1-1239)

QY 5 GluAlaLeuAlaAsp-----GlnThrAspAlaLeuGlnSerGluGluAla 19  
DB 466 GAAGCGGTAGCTGAGGAGGACGCCGCGGAGGAGGATGCCGCGGCTGCCGCC 407

QY 20 AlaValValLysAlaAspAsnAla 27

DB 406 GCTGCTGCTAGGCGCGCACCTGTC 383

RESULT 32

US-09-252-991A-8732/c

; Sequence 8732, Application US/09252991A  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8732  
; LENGTH: 2127  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8732

Alignment Scores:

Pred. No.: 273 Length: 2127  
Score: 52.50 Matches: 14  
Percent Similarity: 71.43% Conservative: 6  
Best Local Similarity: 50.00% Mismatches: 3  
Query Match: 40.70% Indels: 5  
DB: 16 Gaps: 1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-252-991A-8732 (1-2127)

QY 5 GluAlaLeuAlaAsp-----GlnThrAspAlaLeuGlnSerGluGluAla 19  
DB 1608 GAAGCGGTAGCTGAGGAGGACGCCGCGGAGGAGGATGCCGCGGCTGCCGCC 1549

QY 20 AlaValValLysAlaAspAsnAla 27

DB 1548 GCTGCTGCTAGGCGCGCACCTGTC 1525

RESULT 33

US-60-082-302-513/c

; Sequence 513, Application US/60082302  
; GENERAL INFORMATION:  
; APPLICANT: LAGACE, ROBERT E.  
; APPLICANT: CORLEY, NEIL C.

```
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-09-614-150-17530

Alignment Scores:
Pred. No.: 539          Length: 6398
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-614-150-17530 (1-6398)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 2298 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGAGGATGCCGCGCTGCCGCC 2239
Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTCGCGCG 2227

RESULT 26
US-60-191-637-17580/c
; Sequence 17580, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17580
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-637-17580

Alignment Scores:
Pred. No.: 539          Length: 6398
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-191-637-17580 (1-6398)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 2298 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGAGGATGCCGCGCTGCCGCC 2239
Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTCGCGCG 2227

RESULT 27
US-60-191-681-13896/c
; Sequence 13896, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: Li, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000390
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 13896
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-681-13896

Alignment Scores:
Pred. No.: 539          Length: 6398
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-191-681-13896 (1-6398)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 2298 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGAGGATGCCGCGCTGCCGCC 2239
Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTCGCGCG 2227

RESULT 28
US-60-173-464-14400
; Sequence 14400, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14400
; LENGTH: 7203
; TYPE: DNA
; ORGANISM: Drosophila
US-60-173-464-14400

Alignment Scores:
Pred. No.: 618          Length: 7203
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 56 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-173-464-14400 (1-7203)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 4905 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGAGGATGCCGCGCTGCCGCC 4964
Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 4965 GATGCTCGCGCG 4976

RESULT 29
US-60-167-217-17589
; Sequence 17589, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
```

Db 1122 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGCAGAGGATGCCCGCAGCGGCTGCCGCC 1063

```
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Drosophila
US-60-142-845-289

Alignment Scores:
Pred. No.: 124          Length: 1783
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 53 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-142-845-289 (1-1785)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1475 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGATGCGGCGGCTGCCGCC 1534

Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 1535 GATGCTCGGCGG 1546

RESULT 19
US-60-145-134-143
; Sequence 143, Application US/60145134
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; FILE REFERENCE: CL000058
; CURRENT APPLICATION NUMBER: US/60/145,134
; CURRENT FILING DATE: 1999-07-22
; NUMBER OF SEQ ID NOS: 342
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 143
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Drosophila
US-60-145-134-143

Alignment Scores:
Pred. No.: 124          Length: 1785
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 53 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-145-134-143 (1-1785)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1477 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGATGCGGCGGCTGCCGCC 1536

Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 1537 GATGCTCGGCGG 1548

RESULT 20
US-60-173-464-14401/c
; Sequence 14401, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14401
```

```
; LENGTH: 3046
; TYPE: DNA
; ORGANISM: Drosophila
US-60-173-464-14401

Alignment Scores:
Pred. No.: 229          Length: 3046
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 56 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-173-464-14401 (1-3046)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1014 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGATGCGGCGGCTGCCGCC 955

Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 954 GATGCTCGGCGG 943

RESULT 21
US-60-167-217-17590/c
; Sequence 17590, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17590
; LENGTH: 3049
; TYPE: DNA
; ORGANISM: Drosophila
US-60-167-217-17590

Alignment Scores:
Pred. No.: 230          Length: 3049
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 55 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-167-217-17590 (1-3049)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1017 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGATGCGGCGGCTGCCGCC 958

Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 957 GATGCTCGGCGG 946

RESULT 22
US-09-614-150-17531/c
; Sequence 17531, Application US/09614150
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
```

```

CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: 8mm cartridge tape
COMPUTER: SPARC station LX
OPERATING SYSTEM: sunOS4
SOFTWARE: tar
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/621,425
FILING DATE: 25-MARCH-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/561,469
FILING DATE: 17-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,032
FILING DATE: 07-JUNE-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-001CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 3497 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1...3497
US-08-621-425-49

Alignment Scores:
Pred. No.: 83
Score: 57.00
Percent Similarity: 69.57%
Best Local Similarity: 52.17%
Query Match: 44.19%
DB: 10

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-621-425-49 (1-3497)

QY 4 LeuGluaIaIeuAlaAspGlnThrAspAlaIeuGlnSerGluGluAlaAlaValIyIs 23
Db 1528 ATCAAGCCTTTACAGGACAAATTGACGCTTTAGATTCTCAAGAAAGATCGTTAGCAA 1587
QY 24 AlaAspAsn 26
Db 1588 TGGGATAAC 1596

RESULT 18
US-60-142-845-289
: Sequence 289, Application US/60142845
: GENERAL INFORMATION:
: APPLICANT: Kerlavage, Anthony
: TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
: TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
: TITLE OF INVENTION: THEREOF
: FILE REFERENCE: CL000048
: CURRENT APPLICATION NUMBER: US/60/142,845
: CURRENT FILING DATE: 1999-07-08
: NUMBER OF SEQ ID NOS: 704
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 289

```

MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER:  
OPERATING SYSTEM:  
SOFTWARE:  
CURRENT APPLICATION DATA: PCT/US97/19575  
FILING DATE:  
APPLICATION NUMBER: US 08/739,150  
FILING DATE: 28-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/759,739  
FILING DATE: 06-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/891,928  
FILING DATE: 14-JULY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragoras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-001CP10PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1239 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...1239  
PCT-US97-19575-51

Alignment Scores:  
Pred. No.: 25.1 Length: 1239  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x PCT-US97-19575-51 (1-1239)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallys 23  
Db 145 ATACAGCCCTACAGAGCAAAATTCAGCGTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 204  
Qy 24 AlaAspAsn 26  
Db 205 TGGGATAAC 213

RESULT 15  
US-08-759-739-250  
Sequence 250, Application US/08759739  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH ET AL  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI VACCINE COMPOSITIONS THERE  
NUMBER OF SEQUENCES: 608  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER:  
OPERATING SYSTEM:  
SOFTWARE:  
CURRENT APPLICATION DATA: US/08/759,739  
FILING DATE:  
APPLICATION NUMBER: US 08/487,032  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/561,469  
FILING DATE: 17-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/630,405  
FILING DATE: 01-APRIL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/660,742  
FILING DATE: 06-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/736,791  
FILING DATE: 25-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/739,150  
FILING DATE: 28-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragoras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-001CP8  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 250:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1239 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...1239  
US-08-759-739-250

Alignment Scores:  
Pred. No.: 25.1 Length: 1239  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-759-739-250 (1-1239)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallys 23  
Db 145 ATACAGCCCTACAGAGCAAAATTCAGCGTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 204  
Qy 24 AlaAspAsn 26  
Db 205 TGGGATAAC 213

RESULT 16  
US-08-993-002A-4235  
Sequence 4235, Application US/08993002A  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR



```
; APPLICATION NUMBER: US/08/759,739
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APRIL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...666
US-08-759-739-92
Alignment Scores:
Pred. No.: 12.3 Length: 666
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 11 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-759-739-92 (1-666)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 229 ATACAAGCCCTACAGGAGCAAAATTCAGCGTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 288
QY 24 AlaAspAsn 26
Db 289 TGGGATAAC 297
RESULT 13
US-08-993-002A-4234
; Sequence 4234, Application US/08993002A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
```

```
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,002A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4234:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...666
US-08-993-002A-4234
Alignment Scores:
Pred. No.: 12.3 Length: 666
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 13 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-993-002A-4234 (1-666)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 229 ATACAAGCCCTACAGGAGCAAAATTCAGCGTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 288
QY 24 AlaAspAsn 26
Db 289 TGGGATAAC 297
RESULT 14
PCT-US97-19575-51
; Sequence 51, Application PC/TUS9719575
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI AND
; TITLE OF INVENTION: VACCINE COMPOSITIONS THEREOF
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
```

```

RESULT 11
US-08-993-002A-4232
; Sequence 4232, Application US/08993002A
;
; GENERAL INFORMATION:
;
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AN
; TITLE OF INVENTION: RELATING TO HEL
; TITLE OF INVENTION: DIAGNOSTICS AND
;
; NUMBER OF SEQUENCES: 10031
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875

```

; SOFTWARE:

```

1 / APPLICANT: DOUGLAS SMITH ET AL
2 /
3 / TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
4 /
5 / TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI VACCINE COMPOSITIONS THERE
6 /
7 / NUMBER OF SEQUENCES: 608
8 /
9 / CORRESPONDENCE ADDRESS:
10 / ADDRESS: LAHIVE & COCKFIELD
11 / STREET: 60 State Street, Suite 510
12 / CITY: Boston
13 / STATE: Massachusetts
14 / COUNTRY: USA
15 / ZIP: 02109-1875
16 / COMPUTER READABLE FORM:
17 / MEDIUM TYPE: CD-ROM ISO9660
18 / COMPUTER:
19 / OPERATING SYSTEM:
20 / SOFTWARE:
21 / CURRENT APPLICATION DATA:

```

```
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 880
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,032A
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 282:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
;
US-08-487-032A-282

Alignment Scores:
Pred. No.: 10.4 Length: 576
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 8 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-487-032A-282 (1-576)

Qy 4 LeuGluaAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 145 ATACAGCCCTACAGGCAAAATTCAGCTTTAGATTCTCAAGAAAAAGTCGTAGCAAA 204

Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

RESULT 9
US-08-561-469A-282
; Sequence 282, Application US/08561469A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; NUMBER OF SEQUENCES: 994
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/561,469A
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 282:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
;
US-08-561-469A-282

Alignment Scores:
Pred. No.: 10.4 Length: 576
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 9 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-561-469A-282 (1-576)

Qy 4 LeuGluaAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 145 ATACAGCCCTACAGGCAAAATTCAGCTTTAGATTCTCAAGAAAAAGTCGTAGCAAA 204

Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

RESULT 10
US-08-759-739-43
; Sequence 43, Application US/08759739
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI VACCINE COMPOSITIONS THERE
; NUMBER OF SEQUENCES: 608
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,739
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
```

; APPLICANT: Rasmussen, Magnus  
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA 2M BINDING PROTEIN  
; FILE REFERENCE: 100084.415US / N.75312B  
; CURRENT APPLICATION NUMBER: US/09/847,539A  
; CURRENT FILING DATE: 2001-05-01  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 764  
; TYPE: DNA  
; ORGANISM: Streptococcus pyogenes  
US-09-847-539A-27

Alignment Scores:  
Pred. No.: 8,01e-12 Length: 764  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Conservative: 0  
Mismatch: 0  
Query Match: 0  
Indels: 0  
DB: 32 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-27 (1-764)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAala 20  
|||||  
Db 333 TCACATGCTTAGAAGCATTAGCGGATCAAAACACACGCTTTACAAATCAGAAGAAGCTGCG 392  
QY 21 ValVallyAlaAspAsnAlaAala 28  
|||||  
Db 393 GTTGTAAAGCGGATAACGCTGCT 416

## RESULT 5

US-09-847-539A-13  
; Sequence 13, Application US/09847539A  
; GENERAL INFORMATION:  
; APPLICANT: Bjorck, Lars H  
; TITLE OF INVENTION: Rasmussen, Magnus  
; FILE REFERENCE: 100084.415US / N.75312B  
; CURRENT APPLICATION NUMBER: US/09/847,539A  
; CURRENT FILING DATE: 2001-05-01  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 777  
; TYPE: DNA  
; ORGANISM: Streptococcus pyogenes  
US-09-847-539A-13

Alignment Scores:  
Pred. No.: 8,17e-12 Length: 777  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Conservative: 0  
Mismatch: 0  
Query Match: 0  
Indels: 0  
DB: 32 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-13 (1-777)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAala 20  
|||||  
Db 319 AGTCACGCTTAGAAGCATTAGCGGATCAAAACACACGCTTTACAAATCAGAAGAAGCTGCG 378  
QY 21 ValVallyAlaAspAsnAlaAala 28  
|||||  
Db 379 GTTGTAAAGCGGATAACGCTGCT 402

## RESULT 6

US-09-847-539A-15  
; Sequence 15, Application US/09847539A  
; GENERAL INFORMATION:  
; APPLICANT: Bjorck, Lars H  
; APPLICANT: Rasmussen, Magnus

; TITLE OF INVENTION: STREPTOCOCCAL ALPHA 2M BINDING PROTEIN  
; FILE REFERENCE: 100084.415US / N.75312B  
; CURRENT APPLICATION NUMBER: US/09/847,539A  
; CURRENT FILING DATE: 2001-05-01  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 853  
; TYPE: DNA  
; ORGANISM: Streptococcus pyogenes  
US-09-847-539A-15

Alignment Scores:  
Pred. No.: 9,09e-12 Length: 853  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Conservative: 0  
Mismatch: 0  
Query Match: 0  
Indels: 0  
DB: 32 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-15 (1-853)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAala 20  
|||||  
Db 556 AGTCACGCTTAGAAGCATTAGCGGATCAAAACACACGCTTTACAAATCAGAAGAAGCTGCG 615  
QY 21 ValVallyAlaAspAsnAlaAala 28  
|||||  
Db 616 GTTGTAAAGCGGATAACGCTGCT 639

## RESULT 7

US-09-489-039A-318  
; Sequence 318, Application US/09489039A  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709,2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 318  
; LENGTH: 828  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-318

Alignment Scores:  
Pred. No.: 10.7 Length: 828  
Score: 58.00 Matches: 12  
Percent Similarity: 75.00%  
Best Local Similarity: 50.00%  
Conservative: 6  
Mismatch: 6  
Query Match: 44.96%  
Indels: 0  
DB: 18 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-489-039A-318 (1-828)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAalaVal 21  
|||||  
Db 157 GACCGCTGCGAGCGCTCGCGGACACGCTGCAGCGCGCTACGCCATCCAGGCAAGCATC 216  
QY 22 VallyAlaAsp 25  
:::|||||  
Db 217 CTCAAGCGCGAT 228

## RESULT 8

US-08-487-032A-282  
; Sequence 282, Application US/08487032A  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

```
17 57 44.2 3497 10 US-08-621-425-49 Sequence 49, Appl
18 54 41.9 1783 53 US-60-142-845-289 Sequence 289, App
19 54 41.9 1785 53 US-60-145-134-143 Sequence 143, App
20 54 41.9 3046 56 US-60-173-464-14401 Sequence 14401, A
21 54 41.9 3049 55 US-60-167-217-17590 Sequence 17590, A
22 54 41.9 3349 23 US-09-614-150-17531 Sequence 17531, A
23 54 41.9 3349 58 US-60-191-637-17581 Sequence 17581, A
24 54 41.9 3349 58 US-60-191-681-13897 Sequence 13897, A
25 54 41.9 6398 23 US-09-614-150-17530 Sequence 17530, A
26 54 41.9 6398 58 US-60-191-637-17580 Sequence 17580, A
27 54 41.9 6398 58 US-60-191-681-13896 Sequence 13896, A
28 54 41.9 7203 55 US-60-173-464-14400 Sequence 14400, A
29 54 41.9 7206 55 US-60-167-217-17589 Sequence 17589, A
30 54 32275 19 US-09-528-237A-1811 Sequence 1811, Ap
31 52.5 40.7 1239 16 US-09-252-991A-8951 Sequence 8951, Ap
32 52.5 40.7 2127 16 US-09-252-991A-8951 Sequence 8951, Ap
33 52.5 40.7 8830 47 US-60-082-302-513 Sequence 513, App
34 52 40.3 594 22 US-09-565-309A-32821 Sequence 32821, A
35 52 40.3 641 22 US-09-565-309A-32824 Sequence 32824, A
36 52 40.3 939 22 US-09-565-309A-67148 Sequence 67148, A
37 52 40.3 984 28 US-09-708-427-29368 Sequence 29368, A
38 52 40.3 999 22 US-09-565-309A-67147 Sequence 67147, A
39 52 40.3 1230 22 US-09-565-309A-62025 Sequence 62025, A
40 52 40.3 1230 22 US-09-565-329A-545 Sequence 545, App
41 52 40.3 1244 22 US-09-565-309A-52656 Sequence 52656, A
42 52 40.3 16135 23 US-09-614-150-35057 Sequence 35057, A
43 52 40.3 16135 23 US-09-619-049-1340 Sequence 1340, Ap
44 52 40.3 16135 56 US-60-171-627-1946 Sequence 1946, Ap
45 52 40.3 16135 56 US-60-173-464-26803 Sequence 26803, A
```

## ALIGNMENTS

```
RESULT 1
US-09-847-539A-14
; Sequence 14, Application US/09847539A
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
; APPLICANT: Rasmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA ZM BINDING PROTEIN
; FILE REFERENCE: 100084.415US / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-14
```

```
Alignment Scores:
Pred. No.: 4, 57e-12 Length: 469
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 32 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-14 (1-469)

```
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20
|||||
Db 172 TCAGATGCGCTTAGAGCATTAGCGATCAACAGACGCTTTACAAATCAGAAGAGCTGCG 231
|||||
QY 21 ValValLysAlaAspAsnAlaLa 28
|||||
Db 232 GTTGTTAAAGCGGATAACGCTGCT 255
|||||
```

```
RESULT 2
US-09-847-539A-16
; Sequence 16, Application US/09847539A
; GENERAL INFORMATION:
```

```
; APPLICANT: Bjorck, Lars H
; APPLICANT: Rasmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA ZM BINDING PROTEIN
; FILE REFERENCE: 100084.415US / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-16
```

```
Alignment Scores:
Pred. No.: 4, 96e-12 Length: 504
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 32 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-16 (1-504)

```
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20
|||||
Db 208 TCAGATGCGCTTAGAGCATTAGCGATCAACAGACGCTTTACAAATCAGAAGAGCTGCG 267
|||||
QY 21 ValValLysAlaAspAsnAlaLa 28
|||||
Db 268 GTTGTTAAAGCGGATAACGCTGCT 291
|||||
```

```
RESULT 3
US-09-847-539A-12
; Sequence 12, Application US/09847539A
; GENERAL INFORMATION:
```

```
; APPLICANT: Bjorck, Lars H
; APPLICANT: Rasmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA ZM BINDING PROTEIN
; FILE REFERENCE: 100084.415US / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-12
```

```
Alignment Scores:
Pred. No.: 6, 7e-12 Length: 654
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 32 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-12 (1-654)

```
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20
|||||
Db 274 TCAGATGCGCTTAGAGCATTAGCGATCAACAGACGCTTTACAAATCAGAAGAGCTGCG 333
|||||
QY 21 ValValLysAlaAspAsnAlaLa 28
|||||
Db 334 GTTGTTAAAGCGGATAACGCTGCT 357
|||||
```

```
RESULT 4
US-09-847-539A-27
; Sequence 27, Application US/09847539A
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:57:28 ; Search time 558.053 seconds  
(without alignments)  
1085.518 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALEALADTDALQSEEAHVKNADNA 28

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 21979536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n\_model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_093107\_5059/app\_query.fasta.1.526  
-DB=Pending Patents\_NA\_Main -QPMT=fastap -SUFFIX=std.rnmp -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=pct\_THR\_MAX=100  
-THR\_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09847539 -CGN\_1\_1201 -runat\_10102002\_093107\_5059  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAPP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending Patents\_NA\_Main: \*  
1: /cgn2\_6/ptodata/2/pna/PTCUS\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/pna/US06\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/pna/US080\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/pna/US081\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/pna/US082\_COMB.seq: \*  
7: /cgn2\_6/ptodata/2/pna/US083\_COMB.seq: \*  
8: /cgn2\_6/ptodata/2/pna/US084\_COMB.seq: \*  
9: /cgn2\_6/ptodata/2/pna/US085\_COMB.seq: \*  
10: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq: \*  
11: /cgn2\_6/ptodata/2/pna/US087\_COMB.seq: \*  
12: /cgn2\_6/ptodata/2/pna/US088\_COMB.seq: \*  
13: /cgn2\_6/ptodata/2/pna/US089\_COMB.seq: \*  
14: /cgn2\_6/ptodata/2/pna/US090\_COMB.seq: \*  
15: /cgn2\_6/ptodata/2/pna/US091\_COMB.seq: \*  
16: /cgn2\_6/ptodata/2/pna/US092\_COMB.seq: \*  
17: /cgn2\_6/ptodata/2/pna/US093\_COMB.seq: \*  
18: /cgn2\_6/ptodata/2/pna/US094\_COMB.seq: \*  
19: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq: \*  
20: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq: \*  
21: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq: \*  
22: /cgn2\_6/ptodata/2/pna/US095D\_COMB.seq: \*  
23: /cgn2\_6/ptodata/2/pna/US096A\_COMB.seq: \*  
24: /cgn2\_6/ptodata/2/pna/US096B\_COMB.seq: \*  
25: /cgn2\_6/ptodata/2/pna/US096C\_COMB.seq: \*  
26: /cgn2\_6/ptodata/2/pna/US096D\_COMB.seq: \*  
27: /cgn2\_6/ptodata/2/pna/US096E\_COMB.seq: \*  
28: /cgn2\_6/ptodata/2/pna/US097A\_COMB.seq: \*

29: /cgn2\_6/ptodata/2/pna/US097B\_COMB.seq: \*  
30: /cgn2\_6/ptodata/2/pna/US097C\_COMB.seq: \*  
31: /cgn2\_6/ptodata/2/pna/US098A\_COMB.seq: \*  
32: /cgn2\_6/ptodata/2/pna/US098B\_COMB.seq: \*  
33: /cgn2\_6/ptodata/2/pna/US098C\_COMB.seq: \*  
34: /cgn2\_6/ptodata/2/pna/US099A\_COMB.seq: \*  
35: /cgn2\_6/ptodata/2/pna/US099B\_COMB.seq: \*  
36: /cgn2\_6/ptodata/2/pna/US099C\_COMB.seq: \*  
37: /cgn2\_6/ptodata/2/pna/US100\_COMB.seq: \*  
38: /cgn2\_6/ptodata/2/pna/US101\_COMB.seq: \*  
39: /cgn2\_6/ptodata/2/pna/US6000\_COMB.seq: \*  
40: /cgn2\_6/ptodata/2/pna/US6001\_COMB.seq: \*  
41: /cgn2\_6/ptodata/2/pna/US6002\_COMB.seq: \*  
42: /cgn2\_6/ptodata/2/pna/US6003\_COMB.seq: \*  
43: /cgn2\_6/ptodata/2/pna/US6004\_COMB.seq: \*  
44: /cgn2\_6/ptodata/2/pna/US6005\_COMB.seq: \*  
45: /cgn2\_6/ptodata/2/pna/US6006\_COMB.seq: \*  
46: /cgn2\_6/ptodata/2/pna/US6007\_COMB.seq: \*  
47: /cgn2\_6/ptodata/2/pna/US6008\_COMB.seq: \*  
48: /cgn2\_6/ptodata/2/pna/US6009\_COMB.seq: \*  
49: /cgn2\_6/ptodata/2/pna/US6010\_COMB.seq: \*  
50: /cgn2\_6/ptodata/2/pna/US6011\_COMB.seq: \*  
51: /cgn2\_6/ptodata/2/pna/US6012\_COMB.seq: \*  
52: /cgn2\_6/ptodata/2/pna/US6013\_COMB.seq: \*  
53: /cgn2\_6/ptodata/2/pna/US6014\_COMB.seq: \*  
54: /cgn2\_6/ptodata/2/pna/US6015\_COMB.seq: \*  
55: /cgn2\_6/ptodata/2/pna/US6016\_COMB.seq: \*  
56: /cgn2\_6/ptodata/2/pna/US6017\_COMB.seq: \*  
57: /cgn2\_6/ptodata/2/pna/US6018\_COMB.seq: \*  
58: /cgn2\_6/ptodata/2/pna/US6019\_COMB.seq: \*  
59: /cgn2\_6/ptodata/2/pna/US6020\_COMB.seq: \*  
60: /cgn2\_6/ptodata/2/pna/US6021\_COMB.seq: \*  
61: /cgn2\_6/ptodata/2/pna/US6022\_COMB.seq: \*  
62: /cgn2\_6/ptodata/2/pna/US6023\_COMB.seq: \*  
63: /cgn2\_6/ptodata/2/pna/US6024\_COMB.seq: \*  
64: /cgn2\_6/ptodata/2/pna/US6025\_COMB.seq: \*  
65: /cgn2\_6/ptodata/2/pna/US6026\_COMB.seq: \*  
66: /cgn2\_6/ptodata/2/pna/US6027\_COMB.seq: \*  
67: /cgn2\_6/ptodata/2/pna/US6028\_COMB.seq: \*  
68: /cgn2\_6/ptodata/2/pna/US6029\_COMB.seq: \*  
69: /cgn2\_6/ptodata/2/pna/US6030\_COMB.seq: \*  
70: /cgn2\_6/ptodata/2/pna/US6031\_COMB.seq: \*  
71: /cgn2\_6/ptodata/2/pna/US6032\_COMB.seq: \*  
72: /cgn2\_6/ptodata/2/pna/US6033\_COMB.seq: \*  
73: /cgn2\_6/ptodata/2/pna/US6034\_COMB.seq: \*  
74: /cgn2\_6/ptodata/2/pna/US6035\_COMB.seq: \*  
75: /cgn2\_6/ptodata/2/pna/US6036\_COMB.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	129	100.0	469	32	US-09-847-539A-14
2	129	100.0	504	32	US-09-847-539A-16
3	129	100.0	654	32	US-09-847-539A-12
4	129	100.0	764	32	US-09-847-539A-27
5	129	100.0	777	32	US-09-847-539A-13
6	129	100.0	853	32	US-09-847-539A-15
7	58	45.0	828	18	US-09-489-039A-318
8	57	44.2	576	9	US-08-487-032A-282
9	57	44.2	576	9	US-08-561-469A-282
10	57	44.2	576	11	US-08-759-739-43
11	57	44.2	576	13	US-08-993-002A-4232
12	57	44.2	666	11	US-08-759-739-92
13	57	44.2	666	13	US-08-993-002A-4234
14	57	44.2	1239	1	PCT-US97-19575-51
15	57	44.2	1239	11	US-08-759-739-250
16	57	44.2	1239	13	US-08-993-002A-4235

```
Db 1594 GAGCAGAAACTTCAGGATCTTCAAGGAGAAAAGGATGCTTGGATTCTGAAAAAGCAGCAG 1653
Qy 107 ValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAla 126
Db 1654 ATCACTGCACAGAAACAAGACCTGGAGGCAGAGGTGTCCAAGCTGCACAGAGAGAGGTTGCC 1713
Qy 127 LeuAspValLysLysThrLysAspThrLys 136
Db 1714 AAGCTGTCAAAAGAACTAGAGATGCCAAG 1743
```

Search completed: October 13, 2002, 03:34:51  
Job time : 82.8663 secs

APPLICANT: King, Mary-Claire  
 APPLICANT: Lynch, Eric D.  
 APPLICANT: Lee, Ming  
 APPLICANT: Morrow, Jan E.  
 APPLICANT: Welcsh, Piri L.  
 APPLICANT: Leon, Pedro E.  
 TITLE OF INVENTION: Modulators of Actin  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
 STREET: 75 DENISE DRIVE  
 CITY: HILLSBOROUGH  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94010  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/323,735  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 09/080,897  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: OSMAN, RICHARD A.  
 REGISTRATION NUMBER: 36,627  
 REFERENCE/DOCKET NUMBER: UW97-001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650) 343-4341  
 TELEFAX: (650) 343-4342  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4378 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cdna  
 US-09-323-735-3

```

RESULT 40
US-08-899-595-2
; Sequence 2, Application US/0889595
; Patent No. 6111072
; GENERAL INFORMATION:
; APPLICANT: Narumiya, Shuh
; APPLICANT: Takahashi, No. 6111072uakl
; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
; TITLE OF INVENTION: ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,595
; FILING DATE: 24-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-242701
; FILING DATE: 26-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-90170
; FILING DATE: 25-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stephen A. Bent
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 049441/0112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904135
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4399 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 94..3858
; US-08-899-595-2

Alignment Scores:
Pred. No.:          39.8      Length:       4399
Score:             62.00     Matches:        20
Percent Similarity: 40.00%   Conservative:    16
Best Local Similarity: 22.22% Mismatches:         54
Query Match:       7.98%    Indels:           0
DB:                3       Caps:            0

US-09-847-539A-6 (1-159) x US-08-899-595-2 (1-4399)

Qy  47 AspLysGluAlaThrThrAlaIleGluAlaLaSerAspAlaLeuGluAla
|||||              ||||| ::::~::~:
Db 1474 CATAAACAAAGGTGGAAAAATCTGAGGCCCAAGCTACAGAGCTGGAAAAAAGC
Qy  67 AspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLlyAlaAspAsn
|||||              ||||| ::::~::~:
Db 1534 TCAGATTAAACAGCGCCGCACGAGTTACAGACTAGAATAAAGATGGAATAAT
Qy  87 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu

```



```

CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/678,614
FILING DATE: 10-JUL-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-139
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4248 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: RSA 266
FEATURE:
NAME/KEY: ORF
LOCATION: 154..1410
FEATURE:
NAME/KEY: ORF
LOCATION: 1497..3500
US-08-678-614-1

```

APPLICANT: Lee, Ming  
 APPLICANT: Morrow, Jan E.  
 APPLICANT: Welcsh, Piri L.  
 APPLICANT: Leon, Pedro E.  
 TITLE OF INVENTION: Modulators of Actin  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
 STREET: 75 DENISE DRIVE  
 CITY: HILLSBOROUGH  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94010  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/080,897  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: OSMAN, RICHARD A  
 REGISTRATION NUMBER: 36,627  
 REFERENCE/DOCKET NUMBER: UW97-001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650) 343-4341  
 TELEFAX: (650) 343-4342  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4378 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 PS-09-080-897-3

```

; APPLICATION NUMBER: US/09/175,581
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/770,301
; FILING DATE: 20-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1422-287
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-205-8000
; TELEFAX: (703)-205-8050
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3252 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; IDENTIFICATION METHOD: E
US-09-175-581-4
Alignment Scores:
Pred. No.: 26.6 Length: 3252
Score: 62.00 Matches: 22
Percent Similarity: 38.83% Conservative: 18
Best Local Similarity: 21.36% Mismatches: 63
Query Match: 7.98% Indels: 0
DB: 3 Gaps: 0
US-09-847-539A-6 (1-159) x US-09-175-581-4 (1-3252)
QY 2 AspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuLeu 21
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1898 GACGCCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957
QY 22 GlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys 41
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1958 GCGCGTGCTACCTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 2017
QY 42 LysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAspAla 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2018 GAGAGCCCTGCTGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2077
QY 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLys 81
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2078 GCTGAAGCCCAATGGAGGATACCTCCAGCCCTCCAGCCCTCCAGCCCTTAAGGGGATGCA 2137
QY 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2138 GCCCGTGAGGAGAAACACCTCATGATAAAGAAAGTGGGACAAAGTCTGAGGCCCAAG 2197
QY 102 SerGluGlu 104
|||
Db 2198 CCCAATGAG 2206
RESULT 37
US-08-678-614-1
; Sequence 1, Application US/08678614
; Patent No. 6013507
; GENERAL INFORMATION:
; APPLICANT: Tomasz, Alexander
; APPLICANT: Delencastre, Herminia
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS TH
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Floor

```







```
;
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 118..183
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 118..3699
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 4433
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 5195
;
; US-08-317-450B-12
;
Alignment Scores:
Pred. No.: 25.1 Length: 5200
Score: 64.00 Matches: 20
Percent Similarity: 45.16% Conservative: 8
Best Local Similarity: 32.26% Mismatches: 34
Query Match: 8.24% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-317-450B-12 (1-5200)
QY 30 LeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGlu 49
DB 2980 CTCAGAGAGTTTGACCTGCAGGTGGACACACAGAAAGCAGAGCTGAAGAAGCCATGAAG 3039
QY 50 AlaThrThraIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69
DB 3040 AGACTCTCTACATCAGCCAGCAAGAGTTTCAGATGCCAGTGACAAAGACCCAGCAAGCAGAA 3099
QY 70 AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAla 89
DB 3100 AGAGCCCTGGGGAGCGCTCTCTGTCATGCACAGAGCGCAAGAAGTGGCCGGGAGGCC 3159
QY 90 LeuGlu 91
DB 3160 CTGGAA 3165

RESULT 27
US-08-800-593-12
; Sequence 12, Application US/08800593
; Patent No. 6143505
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,593
; FILING DATE: 18-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/317,450
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 94,778-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 12:
```

```
;
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 118..183
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 118..3699
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 4433
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 5195
;
; US-08-800-593-12
;
Alignment Scores:
Pred. No.: 25.1 Length: 5200
Score: 64.00 Matches: 20
Percent Similarity: 45.16% Conservative: 8
Best Local Similarity: 32.26% Mismatches: 34
Query Match: 8.24% Indels: 0
DB: 3 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-800-593-12 (1-5200)
QY 30 LeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGlu 49
DB 2980 CTCAGAGAGTTTGACCTGCAGGTGGACACACAGAAAGCAGAGCTGAAGAAGCCATGAAG 3039
QY 50 AlaThrThraIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69
DB 3040 AGACTCTCTACATCAGCCAGCAAGAGTTTCAGATGCCAGTGACAAAGACCCAGCAAGCAGAA 3099
QY 70 AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAla 89
DB 3100 AGAGCCCTGGGGAGCGCTCTCTGTCATGCACAGAGCGCAAGAAGTGGCCGGGAGGCC 3159
QY 90 LeuGlu 91
DB 3160 CTGGAA 3165

RESULT 28
US-08-961-858-2
; Sequence 2, Application US/08961858
; Patent No. 5834210
; GENERAL INFORMATION:
; APPLICANT: Liu, Shigui
; APPLICANT: Shi, Qinhui
; TITLE OF INVENTION: STABLE TROPONIN SUBUNITS AND COMPLEXES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,858
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
```

Qy 50 AlathrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspClnThr 69  
Db 3040 AGACTCTCTACATGAGGAGGTTTCAGATGCCAGTGCAGAGCCACCAAGCAGAA 3099  
Qy 70 AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAla 89  
Db 3100 AGAGCCCTGGGGAGGCTGCTGCTGATGCACAGAGGCAAGAAATGGGGCGGGAGGCC 3159  
Qy 90 LeuGlu 91  
Db 3160 CTGGAA 3165  
RESULT 25  
US-08-800-593-14  
; Sequence 14, Application US/08800593  
; Patent No. 6143505  
; GENERAL INFORMATION:  
; APPLICANT: Tryggvason, Karl  
; APPLICANT: Kallunki, Pekka  
; APPLICANT: Pyke, Charles  
; TITLE OF INVENTION: Laminin Chains: Diagnostic and  
; TITLE OF INVENTION: Therapeutic Use  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 18-FEB-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/800,593  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Chao, Mark  
; REGISTRATION NUMBER: 37,293  
; REFERENCE/DOCKET NUMBER: 94,778-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4316 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 118..183  
; NAME/KEY: CDS  
; LOCATION: 118..3453  
; NAME/KEY: repeat\_unit  
; LOCATION: 4021..4316  
; OTHER INFORMATION: /rpt\_type= "other"  
; OTHER INFORMATION: /rpt\_family= "HUMAN ALU"  
; FEATURE:  
; NAME/KEY: polyA\_site  
; LOCATION: 4256  
US-08-800-593-14

Alignment Scores:  
Pred. No.: 19.6 Length: 4316  
Score: 64.00 Matches: 20  
Percent Similarity: 45.16% Conservative: 8  
Best Local Similarity: 32.26% Mismatches: 34  
Query Match: 8.24% Indels: 0  
DB: 3 Gaps: 0  
US-09-847-539A-6 (1-159) x US-08-800-593-14 (1-4316)  
Qy 30 LeuArgAsnGluGluArgAlaIleAspClnLeuLysLysGlnAlaIleGluAspLysGlu 49  
Db 2980 CTCAGAGAGTTTGACCTGCAGTGGACAAAGAAAGCAGAGCTGAAGAACCCATGAAG 3039  
Qy 50 AlathrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspClnThr 69  
Db 3040 AGACTCTCTACATGAGGAGGTTTCAGATGCCAGTGCACAGAGCCACCAAGCAGAA 3099  
Qy 70 AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAla 89  
Db 3100 AGAGCCCTGGGGAGGCTGCTGCTGATGCACAGAGGCAAGAAATGGGGCGGGAGGCC 3159  
Qy 90 LeuGlu 91  
Db 3160 CTGGAA 3165  
RESULT 26  
US-08-317-450B-12  
; Sequence 12, Application US/08317450B  
; Patent No. 5660982  
; GENERAL INFORMATION:  
; APPLICANT: Tryggvason, Karl  
; APPLICANT: Kallunki, Pekka  
; APPLICANT: Pyke, Charles  
; TITLE OF INVENTION: Laminin Chains: Diagnostic and  
; TITLE OF INVENTION: Therapeutic Use  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BANNER & ALLEGRETTI, LTD.  
; STREET: Ten South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,450B  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Chao, Mark  
; REGISTRATION NUMBER: 37,293  
; REFERENCE/DOCKET NUMBER: 94,778  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 118..183  
; NAME/KEY: CDS  
; LOCATION: 118..3699









```
;
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 222..425
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 451..747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747..1109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1109..2014
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2034..2747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2747..3109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3109..3444
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3444..3728
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3731..4855
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function= "potential coding
; OTHER INFORMATION: sequence"
; OTHER INFORMATION: /product= "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
```

```
;
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; US-08-402-282-3
; Alignment Scores:
; Pred. No.: 2.66e+03 Length: 15664
; Score: 42.00 Matches: 8
; Percent Similarity: 63.64% Conservative: 6
; Best Local Similarity: 36.36% Mismatches: 8
; Query Match: 32.56% Indels: 0
; Db: 1 Gaps: 0
; US-09-847-539A-6_COPY_59_86 (1-28) x US-08-402-282-3 (1-15664)
; Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
; Db 5643 GAGGGCGTGGCCACATCGAGGACNAGCTCGACGTGTTCAGCAGACACGCGCCGCG 5702
```



; CITY: Chicago  
; STATE: IL  
; COUNTRY: US  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release. #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/581,148C  
; FILING DATE: 29-DEC-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Larcher, Carol  
; REGISTRATION NUMBER: 35243  
; REFERENCE/DOCKET NUMBER: 71380  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 616-5600  
; TELEFAX: (312) 616-5700  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1903 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1665  
; US-08-581-148C-15

Alignment Scores:  
Pred. No.: 195 Length: 1903  
Score: 42.00 Matches: 9  
Percent Similarity: 60.00% Conservative: 6  
Best Local Similarity: 36.00% Mismatches: 10  
Query Match: 32.56% Indels: 0  
DB: 3 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-581-148C-15 (1-1903)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21  
::: |||||::: ||| ||| :::  
Db 918 GAAATACCTGGAAGCCCTAGCGGACCGCTCCAGGCTCGTGGAGGAGCGCGGAGCTT 859  
QY 22 ValLysAlaAspAsn 26  
::: |||||:::  
Db 858 ATAGAAGCTGCAGT 844

RESULT 13

US-08-285-641-20  
; Sequence 20, Application US/08285641  
; Patent No. 5726299  
; GENERAL INFORMATION:  
; APPLICANT: Zupancic, Thomas J.  
; APPLICANT: Yukawa, Hideaki  
; TITLE OF INVENTION: PROMOTER DNA FRAGMENT FROM CORNYNEFORM BACTERIA  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE:  
; COMPUTER:  
; OPERATING SYSTEM:  
; SOFTWARE:  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/285,641

; FILING DATE: 1-AUGUST-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: 08/076,091  
; FILING DATE: 15-JUN-1993  
; APPLICATION NUMBER: US 07/709,151  
; FILING DATE: 29-MAY-1991  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE:  
; TELEFAX:  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2213 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Brevibacterium flavum  
; STRAIN: MJ-233  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: 1-2213  
; IDENTIFICATION METHOD: experiment  
; US-08-285-641-20  
Alignment Scores:  
Pred. No.: 235 Length: 2213  
Score: 42.00 Matches: 9  
Percent Similarity: 70.59% Conservative: 3  
Best Local Similarity: 52.94% Mismatches: 5  
Query Match: 32.56% Indels: 0  
DB: 1 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-285-641-20 (1-2213)  
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 17  
|||||::: ||| :::  
Db 1388 TCGGATGCTCTCAGGACGCCGCTGAGCAAGCAGATGCTGCTGAATTCGA 1438

RESULT 14

US-08-285-641-16/c  
; Sequence 16, Application US/08285641  
; Patent No. 5726299  
; GENERAL INFORMATION:  
; APPLICANT: Zupancic, Thomas J.  
; APPLICANT: Yukawa, Hideaki  
; TITLE OF INVENTION: PROMOTER DNA FRAGMENT FROM CORNYNEFORM BACTERIA  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE:  
; COMPUTER:  
; OPERATING SYSTEM:  
; SOFTWARE:  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/285,641  
; FILING DATE: 1-AUGUST-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER:  
; APPLICATION NUMBER: 08/076,091  
; FILING DATE: 15-JUN-1993  
; APPLICATION NUMBER: US 07/709,151  
; FILING DATE: 29-MAY-1991  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE:



```
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1221 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; Alignment Scores:
; Pred. NO.: 74.5 Length: 1221
; Score: 43.00 Matches: 9
; Percent Similarity: 51.85% Conservativeness: 5
; Best Local Similarity: 33.33% Mismatches: 13
; Query Match: 33.33% Indels: 0
; DB: 4 Gaps: 0
; US-09-847-539A-6_COPY_59_86 (1-28) x US-09-238-557-3 (1-1221)
;
; Alignment Scores:
; Pred. NO.: 74.5 Length: 1221
; Score: 43.00 Matches: 9
; Percent Similarity: 51.85% Conservativeness: 5
; Best Local Similarity: 33.33% Mismatches: 13
; Query Match: 33.33% Indels: 0
; DB: 4 Gaps: 0
; US-09-847-539A-6_COPY_59_86 (1-28) x US-09-238-557-3 (1-1221)
;
; Qy 2 AspalalaLeuAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
; Db 499 GACCATCAAGAACCTTGGGTGATAGTCTAGAGCAATTGCAGAGCAGAGAAAGCTGGTATT 558
; Qy 22 VallysAlaAspAsnAlaAla 28
; Db 559 TTCAAGGCTGCTAAGAGGCA 579
;
; RESULT 8
; US-08-934-846-1
; Sequence 1, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folic
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/934,846
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-238-557-1
;
; Alignment Scores:
; Pred. NO.: 77 Length: 1254
; Score: 43.00 Matches: 9
; Percent Similarity: 51.85% Conservativeness: 5
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-934-846-1
;
; Alignment Scores:
; Pred. NO.: 77 Length: 1254
; Score: 43.00 Matches: 9
; Percent Similarity: 51.85% Conservativeness: 5
;
; US-09-847-539A-6_COPY_59_86 (1-28) x US-08-934-846-1 (1-1254)
;
; Qy 2 AspalalaLeuAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
; Db 499 GACCATCAAGAACCTTGGGTGATAGTCTAGAGCAATTGCAGAGCAGAGAAAGCTGGTATT 558
; Qy 22 VallysAlaAspAsnAlaAla 28
; Db 559 TTCAAGGCTGCTAAGAGGCA 579
;
; RESULT 9
; US-09-238-557-1
; Sequence 1, Application US/09238557
; Patent No. 6163472
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folic
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/238,557
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-238-557-1
;
; Alignment Scores:
; Pred. NO.: 77 Length: 1254
; Score: 43.00 Matches: 9
; Percent Similarity: 51.85% Conservativeness: 5
```

```
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,022
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1772 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-960-022-13

Alignment Scores:
Pred. No.: 78.4 Length: 1772
Score: 44.00 Matches: 9
Percent Similarity: 68.00% Conservative: 8
Best Local Similarity: 36.00% Mismatches: 8
Query Match: 34.11% Indels: 0
DB: 2 Gaps: 0

US-08-847-539A-6_COPY_59_86 (1-28) x US-08-960-022-13 (1-1772)
QY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaValVal 22
Db 795 AGCATCAAGGCACAAAGGCGACGACGCTGCTGAAGGAGAGGAGGCGCGCTTGTG 854
QY 23 LysAlaAspAsnAla 27
Db 855 GAAAGTAACACGCA 869

RESULT 6
US-08-934-846-3
; Sequence 3, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folc
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,557
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
```

```
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,846
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1221 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-934-846-3

Alignment Scores:
Pred. No.: 74.5 Length: 1221
Score: 43.00 Matches: 9
Percent Similarity: 51.85% Conservative: 5
Best Local Similarity: 33.33% Mismatches: 13
Query Match: 33.33% Indels: 0
DB: 2 Gaps: 0

US-08-847-539A-6_COPY_59_86 (1-28) x US-08-934-846-3 (1-1221)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaValVal 21
Db 499 GACCATCAAGAAACCTTGGGTGATCTCTAGAAGCAATTGCACAGCAGAAAGCTGTATT 558
QY 22 ValLysAlaAspAsnAlaAla 28
Db 559 TTCAAGGCTGGTAAAGGCA 579

RESULT 7
US-09-238-557-3
; Sequence 3, Application US/09238557
; Patent No. 6165472
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folc
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,557
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
```



; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Alignment Scores:  
Pred. No.: 2.94e+05 Length: 4403765  
Score: 47.00 Matches: 9  
Percent Similarity: 54.55% Conservative: 3  
Best Local Similarity: 40.91% Mismatches: 10  
Query Match: 36.43% Indels: 0  
DB: 4 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-103-840A-2 (1-4403765)

Qy 7 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsn 26  
Db 4036256 TTGATGGACGCCCGACCTGCTGGAGGCGAAGTAACCATCGATATCGACAA 4036197

Qy 27 AlaAla 28  
Db 4036196 GGTGCT 4036191

RESULT 3  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Alignment Scores:  
Pred. No.: 2.94e+05 Length: 4411529  
Score: 47.00 Matches: 9  
Percent Similarity: 54.55% Conservative: 3  
Best Local Similarity: 40.91% Mismatches: 10  
Query Match: 36.43% Indels: 0  
DB: 4 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-103-840A-1 (1-4411529)

Qy 7 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsn 26  
Db 4044185 TTGATGGACGCCCGACCTGCTGGAGGCGAAGTAACCATCGATATCGACAA 4044126

Qy 27 AlaAla 28  
Db 4044125 GGTGCT 4044120

RESULT 4  
US-08-961-083-191  
; Sequence 191, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,083  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 191:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1033 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-083-191

Alignment Scores:  
Pred. No.: 40.2 Length: 1033  
Score: 44.00 Matches: 11  
Percent Similarity: 65.00% Conservative: 2  
Best Local Similarity: 55.00% Mismatches: 7  
Query Match: 34.11% Indels: 0  
DB: 3 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-961-083-191 (1-1033)

Qy 8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 27  
Db 8 GCAGGGCAGACAGATGCTCCCAATTGAAAGGCGGCGAGTTAGCCACAGAGGAGAAAGCA 67

RESULT 5  
US-08-960-022-13  
; Sequence 13, Application US/08960022  
; Patent No. 5976837  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:17:36 : Search time 9.13369 Seconds  
(without alignments)  
753.008 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALEALADQTDALQSEEAAYVKADNAA 28

Scoring table:

BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09847539/runat\_10102002\_092549\_3456/app\_query.fasta\_1.526  
-DB=Issued\_Patents\_NA -QFMT=Fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09847539.ecgn\_1\_13 @runat\_10102002\_092549\_3456 -NCPUS=6 -ICPU=3  
-NO\_XUPXY -NO\_MMAP -LARGEOVERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DSELP=6 -DELEXT=7

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/lna/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/lna/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/lna/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/lna/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	47	36.4	59065	4	US-09-813-817-3
C 2	47	36.4	4403765	4	US-09-103-840A-2
C 3	47	36.4	4411529	4	US-09-103-840A-1
C 4	44	34.1	1033	3	US-08-961-083-191
C 5	44	34.1	1772	2	US-08-960-022-13
C 6	43	33.3	1221	2	US-08-934-846-3
C 7	43	33.3	1221	4	US-09-238-557-3
C 8	43	33.3	1254	2	US-08-934-846-1
C 9	43	33.3	1254	4	US-09-238-557-1
C 10	43	33.3	4224	1	US-08-612-521-1
C 11	42	32.6	621	2	US-08-943-915-4
C 12	42	32.6	1903	3	US-08-581-148C-15

13	42	32.6	2213	1	US-08-285-641-20	Sequence 20, Appl
14	42	32.6	2248	1	US-08-285-641-16	Sequence 16, Appl
15	42	32.6	4447	4	US-09-521-668B-17	Sequence 17, Appl
16	42	32.6	4447	4	US-09-521-668B-19	Sequence 19, Appl
17	42	32.6	15664	1	US-08-402-282-3	Sequence 3, Appl1
18	42	32.6	15664	1	US-08-508-004-3	Sequence 3, Appl1
19	42	32.6	15664	1	US-08-402-066-3	Sequence 3, Appl1
20	42	32.6	15664	1	US-08-402-068-3	Sequence 3, Appl1
21	42	32.6	68750	3	US-09-335-409-1	Sequence 1, Appl1
22	42	32.6	68750	4	US-09-568-102-1	Sequence 1, Appl1
23	42	32.6	68750	4	US-09-567-969-1	Sequence 1, Appl1
24	42	32.6	68750	4	US-09-568-480-1	Sequence 1, Appl1
25	42	32.6	68750	4	US-09-568-486-1	Sequence 1, Appl1
26	42	32.6	68750	4	US-09-568-472-1	Sequence 1, Appl1
27	41	31.8	1129	5	PCT-US93-01676A-3	Sequence 3, Appl1
28	41	31.8	1185	1	US-07-876-280-8	Sequence 8, Appl1
29	41	31.8	1185	1	US-08-049-783-5	Sequence 5, Appl1
30	41	31.8	1185	1	US-08-316-301A-9	Sequence 9, Appl1
31	41	31.8	1185	2	US-08-904-278-9	Sequence 9, Appl1
32	41	31.8	1185	4	US-09-076-137-9	Sequence 9, Appl1
33	41	31.8	1185	4	US-09-222-594-9	Sequence 9, Appl1
34	41	31.8	1185	5	PCT-US92-03624-9	Sequence 9, Appl1
35	41	31.8	1242	1	US-08-252-966B-13	Sequence 13, Appl
36	41	31.8	1680	1	US-08-140-728A-2	Sequence 2, Appl1
37	41	31.8	1680	1	US-08-546-666-2	Sequence 2, Appl1
38	41	31.8	1680	2	US-08-916-745-2	Sequence 2, Appl1
39	41	31.8	1680	2	US-09-042-929-2	Sequence 2, Appl1
40	41	31.8	1680	2	US-08-546-661-2	Sequence 2, Appl1
41	41	31.8	1680	2	US-09-042-960-2	Sequence 2, Appl1
42	41	31.8	1680	3	US-09-198-650-2	Sequence 2, Appl1
43	41	31.8	1680	3	US-09-042-913-2	Sequence 2, Appl1
44	41	31.8	1680	3	US-09-042-937-2	Sequence 2, Appl1
45	41	31.8	12565	4	US-09-345-217-3	Sequence 3, Appl1

#### ALIGNMENTS

##### RESULT 1

US-09-813-817-3/c  
; Sequence 3, Application US/09813817  
; Patent No. 6340583

; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al.

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; FILE REFERENCE: CL001178

; CURRENT FILING DATE: 2001-03-22

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatsSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 59065

; TYPE: DNA

; ORGANISM: Human

US-09-813-817-3

Alignment Scores:  
Pred. No.: 1.77e+03 Length: 59065  
Score: 47.00 Matches: 10  
Percent Similarity: 76.47% Conservative: 3  
Best Local Similarity: 58.82% Mismatches: 4  
Query Match: 36.43% Indels: 0  
DB: 4 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-09-813-817-3 (1-59065)

QY 8 AlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValysala 24

DB 21239 GCTGACCAACTGACTCTTAGAACACGAGTGTGACGACAGTCTTCAAGCG 21189

##### RESULT 2

US-09-103-840A-2/c

;; FILING DATE:  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Sutton, Jeffrey A.  
;; REGISTRATION NUMBER: 34028  
;; REFERENCE/DOCKET NUMBER: P50164  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (215) 270-5024  
;; TELEFAX: (215) 270-5090  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 58 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: circular  
;; MOLECULE TYPE: peptide  
;; PCT-US94-06655-12

Query Match 30.2%; Score 39; DB 5; Length 58;  
Best Local Similarity 56.2%; Pred. No. 28;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 6 ALADQTDALQSEAAV 21  
  II : IIII : II :  
Db 39 ALNQEVDALQEEVAAL 54

RESULT 39  
US-08-961-858-6  
; Sequence 6, Application US/08961858  
; Patent No. 5834210  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Shigui  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 1112-1-044 CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 288 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
US-08-961-858-6

Query Match 30.2%; Score 39; DB 2; Length 288;  
Best Local Similarity 32.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

Qy 4 LEALADQTDALQSEAAVVKADNAA 28  
  : I : : : : IIII : : II  
Db 4 IEEVVEEYEEEEEAEAAVEQEAA 28

RESULT 40  
US-09-089-593-6  
; Sequence 6, Application US/09089593  
; Patent No. 6060278  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Shigui  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 1112-1-044 CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 288 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
US-09-089-593-6

Query Match 30.2%; Score 39; DB 3; Length 288;  
Best Local Similarity 32.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

Qy 4 LEALADQTDALQSEAAVVKADNAA 28  
  : I : : : : IIII : : II  
Db 4 IEEVVEEYEEEEEAEAAVEQEAA 28

Search completed: October 13, 2002, 02:12:51  
Job time : 6.34225 secs



;; TITLE OF INVENTION: Mad or Max  
;; NUMBER OF SEQUENCES: 19  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Christensen, O'Connor, Johnson, and KindnessPLLC  
;; STREET: 1420 Fifth Ave., Suite 2600  
;; CITY: Seattle  
;; STATE: Washington  
;; COUNTRY: USA  
;; ZIP: 98101-2347  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/252.966B  
;; FILING DATE: 01-JUN-1994  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Shelton, Dennis K.  
;; REGISTRATION NUMBER: 26,997.  
;; REFERENCE/DOCKET NUMBER: FHCRI7694  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (206) 682-8100  
;; TELEFAX: (206) 224-0779  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1253 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; DESCRIPTION: translation of maina cdna; see Figure 23  
;; HYPOTHEICAL: YES  
;; ORIGINAL SOURCE:  
;; ORGANISM: Mus musculus  
;; US-08-252-966B-12

Query Match 31.0%; Score 40; DB 1; Length 1253;  
Best Local Similarity 39.1%; Pred. No. 7.6e+02;  
Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 LEALADQTDALQSEEAHVVKADN 26  
||| : | : || : : ||  
Db 652 LEATQKLSRLSAEQAKFRLDN 674  
RESULT 33  
US-07-920-281C-3  
; Sequence 3, Application US/07920281C  
; Patent No. 5739026  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; APPLICANT: Liljestrom, Peter  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/920.281C  
; FILING DATE: 13-AUG-1992  
; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Murphy Jr., Gerald M.  
;; REGISTRATION NUMBER: 28,977  
;; REFERENCE/DOCKET NUMBER: 828-103P  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 703-241-1300  
;; TELEFAX: 703-241-2848  
;; TELEX: 248345  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1253 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-07-920-281C-3  
Query Match 31.0%; Score 40; DB 1; Length 1253;  
Best Local Similarity 33.3%; Pred. No. 7.6e+02;  
Matches 8; Conservative 6; Mismatches 10; Indels 0; Gaps 0;  
QY 5 EALADQTDALQSEEAHVVKADNA 28  
||| : | : || : : ||  
Db 920 EAYVDRSDVCRHDHASAYKAHTAS 943  
RESULT 34  
US-08-466-277-3  
; Sequence 3, Application US/08466277  
; Patent No. 6190666  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; APPLICANT: Liljestrom, Peter  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466.277  
; FILING DATE: 06-Jun-1995  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/920.281  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 828-103P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-241-1300  
; TELEFAX: 703-241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1253 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-08-466-277-3  
Query Match 31.0%; Score 40; DB 4; Length 1253;  
Best Local Similarity 33.3%; Pred. No. 7.6e+02;



```
QY      5 EALADQTDALQSEAAVVKADN 26
      | | : | | : | | : | | : |
Db     231 EELKKVDDLLKKELEAAIRAE 252

RESULT 26
US-08-858-207A-394
; Sequence 394, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328e1 Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glimi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 394:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 119 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-394

Query Match      31.0%; Score 40; DB 4; Length 119;
Best Local Similarity 30.8%; Pred. No. 47;
Matches 8; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY      2 DALEALADQTDALQSEAAVVKADNA 27
      | : | | | | | | | | : | : | : |
Db      8 DKVDALLDQLSANSKDKTALLESTKA 33

RESULT 27
US-08-457-245-3
; Sequence 3, Application US/08457245
; Patent No. 5573915
; GENERAL INFORMATION:
; APPLICANT: BARRY III, Clifton E.
; APPLICANT: YUAN, Ying
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA INVOLVED
; IN THE BIOSYNTHESIS OF CYCLOPROPANATED MYCOLIC ACIDS IN
; MYCOBACTERIUM TUBERCULOSIS
; TITLE OF INVENTION:

Query Match      31.0%; Score 40; DB 4; Length 549;
Best Local Similarity 47.6%; Pred. No. 2.9e+02;
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,245
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Chambers, Guy W.
; REGISTRATION NUMBER: 30,617
; REFERENCE/DOCKET NUMBER: 15280-216000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: MACE protein
US-08-457-245-3

Query Match      31.0%; Score 40; DB 1; Length 287;
Best Local Similarity 61.1%; Pred. No. 1.3e+02;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY      1 SDALALADQTDALQSEE 18
      | | | | | | | | | | | |
Db     240 SAALQANKQAIALQSEE 257

RESULT 28
US-09-245-041-9
; Sequence 9, Application US/09245041
; Patent No. 6274339
; GENERAL INFORMATION:
; APPLICANT: Moore, K.
; APPLICANT: Nagle, D.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
; OF BODY WEIGHT DISORDERS INCLUDING OBESITY
; FILE REFERENCE: 7853-136
; CURRENT APPLICATION NUMBER: US/09/245,041
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/093,630
; EARLIER FILING DATE: 1998-07-21
; EARLIER APPLICATION NUMBER: 60/104,978
; EARLIER FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: SITE
; LOCATION: all Xaa positions
; OTHER INFORMATION: Xaa-unknown amino acid
US-09-245-041-9

Query Match      31.0%; Score 40; DB 4; Length 549;
Best Local Similarity 47.6%; Pred. No. 2.9e+02;
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
```

;; TITLE OF INVENTION: NO. 6166195el Nematode-Active Toxins and Genes Which Code  
;; FILE OF INVENTION: Therefor  
;; CURRENT APPLICATION NUMBER: MA-20CCCD2  
;; CURRENT FILING DATE: 1998-05-12  
;; EARLIER APPLICATION NUMBER: 08/316,301  
;; EARLIER FILING DATE: 1994-09-30  
;; NUMBER OF SEQ ID NOS: 42  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 10  
;; LENGTH: 395  
;; TYPE: PRT  
;; ORGANISM: Bacillus thuringiensis  
US-09-076-137-10

Query Match 31.8%; Score 41; DB 4; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.4e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

OY 5 EALADQTDALQSEAAVVKADN 26  
DB 231 EELLKKVDDLKKELEAAIKAE 252

## RESULT 24

US-09-222-594-10  
; Sequence 10, Application US/09222594  
; Patent No. 6303364

## ; GENERAL INFORMATION:

; APPLICANT: Thompson, Mark  
; APPLICANT: Knuth, Mark  
; APPLICANT: Cardineau, Guy

; TITLE OF INVENTION: Bacillus thuringiensis Toxins with Improved

; TITLE OF INVENTION: Activity

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Saliwanchik & Saliwanchik

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: Florida

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/222,594

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/904,278

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Sanders, Jay M.

; REGISTRATION NUMBER: 39,355

; REFERENCE/DOCKET NUMBER: MA-702

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (352) 375-8100

; TELEFAX: (352) 372-5800

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 395 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: BACILLUS THURINGIENSIS

; INDIVIDUAL ISOLATE: PS69D1

;; IMMEDIATE SOURCE:  
;; CLONE: E. coli NM522(pMYC2317) NRRL B-18816  
;; FEATURE:  
;; NAME/KEY: Protein  
;; LOCATION: 1..395  
US-09-222-594-10

Query Match 31.8%; Score 41; DB 4; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.4e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

OY 5 EALADQTDALQSEAAVVKADN 26  
DB 231 EELLKKVDDLKKELEAAIKAE 252

## RESULT 25

PCT-US92-03624-10

; Sequence 10, Application PC/TUS9203624

; GENERAL INFORMATION:

; APPLICANT: Schnepf, Harry E.

; APPLICANT: Schwab, George E.

; APPLICANT: Payne, Jewel M.

; APPLICANT: Narva, Kenneth E.

; APPLICANT: Poncerrada, Luis

; TITLE OF INVENTION: Novel Nematode-Active Toxins and Genes

; TITLE OF INVENTION: Which Code Therefor

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David R. Saliwanchik

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/03624

; FILING DATE: 19920501

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Saliwanchik, David R.

; REGISTRATION NUMBER: 31,794

; REFERENCE/DOCKET NUMBER: MA20C2C1C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 904-375-8100

; TELEFAX: 904-372-5800

; INFORMATION FOR SEQ ID NO: 10 (PS69D1):

; SEQUENCE CHARACTERISTICS:

; LENGTH: 395 amino acids

; TYPE: AMINO ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: BACILLUS THURINGIENSIS

; INDIVIDUAL ISOLATE: PS69D1

; IMMEDIATE SOURCE:

; CLONE: E. coli NM522(pMYC2317) NRRL B-18816

; FEATURE:

; NAME/KEY: Protein

; LOCATION: 1..395

; PCT-US92-03624-10

Query Match 31.8%; Score 41; DB 5; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.4e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;





; APPLICANT: Foncerrada, Luis

QY	2	DALEALADQTDALQSEEA	VVKADNAA	28
		: :	:        :	
Db	3434	DALRMAEAGAEOQIVEAD	VARRDDVA	3460

```

RESULT 15
US-09-568-480-6
; Sequence 6, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Lignon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-3082A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: prt
; ORGANISM: Sorangium cellulosum
US-09-568-480-6

```

```

RESULT 16
US-09-568-486-6
; Sequence 6, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-486-6

```

```

RESULT 17
US-09-568-472-6
; Sequence 6, Application US/09568472
; Patent NO. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-6

```

```

1 RESULT 18
2 PCT-US93-01676A-4
3 : Sequence 4, Application PC/TUS9301676A
4 : GENERAL INFORMATION:
5 : APPLICANT: Stambrook, Peter J.
6 : TITLE OF INVENTION: MUTAGENICITY TESTING USING REPORTER
7 : TITLE OF INVENTION: GENES WITH MODIFIED METHYLATION FREQUENCIES
8 : NUMBER OF SEQUENCES: 8
9 : CORRESPONDENCE ADDRESS:
10 : ADDRESSEE: Broadway and Neimark
11 : STREET: 419 Seventh Street, N.W., Suite 300
12 : CITY: Washington
13 : STATE: D.C.
14 : COUNTRY: USA
15 : ZIP: 20004
16 : COMPUTER READABLE FORM:
17 : MEDIUM TYPE: Floppy disk
18 : COMPUTER: IBM PC compatible
19 : OPERATING SYSTEM: PC-DOS/MS-DOS
20 : SOFTWARE: Patent In Release #1.0, Version #1.25
21 : CURRENT APPLICATION DATA:
22 : APPLICATION NUMBER: PCT/US93/01676A
23 : FILING DATE: 19930226
24 : CLASSIFICATION:
25 : PRIORITY APPLICATION DATA:
26 : APPLICATION NUMBER: 07/842,644
27 : FILING DATE: 02-FEB-1992
28 : ATTORNEY/AGENT INFORMATION:
29 : NAME: Cooper, Iver P.
30 : REGISTRATION NUMBER: 28,005
31 : REFERENCE/DOCKET NUMBER: STAMBROOK 1
32 : TELECOMMUNICATION INFORMATION:
33 : TELEPHONE: 202-628-5197
34 : TELEFAX: 202-737-3528
35 : TELEX: 248633
36 : INFORMATION FOR SEQ ID NO: 4:
37 : SEQUENCE CHARACTERISTICS:
38 : LENGTH: 370 amino acids
39 : TYPE: AMINO ACID
40 :

```

; TYPE: PRT  
 ; ORGANISM: Sorangium cellulosum  
 US-09-335-409-6

Query Match 32.6%; Score 42; DB 4; Length 3798;  
Best Local Similarity 37.0%; Pred. No. 1.4e+03;  
Matches 10; Conservative 6; Mismatches 11; Indels

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,100
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,064
; FILING DATE:
; APPLICATION NUMBER: US 08/032,235
; FILING DATE: 17-MAR-1993
; APPLICATION NUMBER: GB 9206016.9
; FILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5773700ris, Allen E.
; REGISTRATION NUMBER: 34,490
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-764-100-25

Query Match 33.3%; Score 43; DB 1; Length 744;
Best Local Similarity 38.5%; Pred. No. 1.5e+02;
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 SDALEALADQTDALQSEEAHVADN 26
   1: 11 : : : : : : : : : :
Db 432 SERLEPRLEALEASERVALLEASN 457

RESULT 8
US-08-710-749-8
; Sequence 8, Application US/08710749
; Patent No. 5955089
; GENERAL INFORMATION:
; APPLICANT: Brilles, David E.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Becker, Robert.
; TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtlis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,749
; FILING DATE: 20-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2074
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
```

```
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 amino acids
; TYPE: amino acid
; STRANDEDNESS: n/a
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
; US-08-710-749-8

Query Match 32.6%; Score 42; DB 2; Length 102;
Best Local Similarity 45.5%; Pred. No. 20;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 4 LEALADQTDALQSEEAHVAKD 25
   1: 1:1: 1:1: 1:1: 1:1:
Db 36 LEELSKIDELDAEIAKNLAKD 57
   1: 1:1: 1:1: 1:1: 1:1:

RESULT 9
US-09-521-668B-18
; Sequence 18, Application US/09521668B
; Patent No. 6303383
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, JUN
; APPLICANT: KANNO, SOHEI
; APPLICANT: KIMURA, EIICHIRO
; APPLICANT: MATSUI, KAHIKO
; APPLICANT: NAKAMATSU, TSUYOSHI
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA
; FILE REFERENCE: 0010-1093-0
; CURRENT APPLICATION NUMBER: US/09/521,668B
; CURRENT FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: JP 11-69896
; PRIOR FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum
; US-09-521-668B-18

Query Match 32.6%; Score 42; DB 4; Length 427;
Best Local Similarity 47.4%; Pred. No. 1.1e+02;
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 LEALADQTDALQSEEAHV 22
   1: 111111: 1:1:
Db 256 LDMLADQSDAGMDAVLV 274

RESULT 10
US-09-521-668B-20
; Sequence 20, Application US/09521668B
; Patent No. 6303383
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, JUN
; APPLICANT: KANNO, SOHEI
; APPLICANT: KIMURA, EIICHIRO
; APPLICANT: MATSUI, KAHIKO
; APPLICANT: NAKAMATSU, TSUYOSHI
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA
; FILE REFERENCE: 0010-1093-0
; CURRENT APPLICATION NUMBER: US/09/521,668B
; CURRENT FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: JP 11-69896
; PRIOR FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum
```

```
;
;
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-238-557-4
;
Query Match 33.3%; Score 43; DB 4; Length 407;
Best Local Similarity 33.3%; Pred. No. 72;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 2 DALEALADQTDALQSEEAHVKNAA 28
Db 167 DHQETLGDLSLEIAEQKAGIFKAGKKA 193

RESULT 5
US-08-934-846-2
; Sequence 2, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL FOLC
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,846
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-238-557-2
;
Query Match 33.3%; Score 43; DB 2; Length 418;
Best Local Similarity 33.3%; Pred. No. 74;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 2 DALEALADQTDALQSEEAHVKNAA 28
Db 167 DHQETLGDLSLEIAEQKAGIFKAGKKA 193

RESULT 6
US-09-238-557-2
;
Query Match 33.3%; Score 43; DB 2; Length 418;
Best Local Similarity 33.3%; Pred. No. 74;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 2 DALEALADQTDALQSEEAHVKNAA 28
Db 167 DHQETLGDLSLEIAEQKAGIFKAGKKA 193
```

```
;
;
; Sequence 2, Application US/09238557
; Patent No. 6165472
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL FOLC
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,557
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-238-557-2
;
Query Match 33.3%; Score 43; DB 4; Length 418;
Best Local Similarity 33.3%; Pred. No. 74;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 2 DALEALADQTDALQSEEAHVKNAA 28
Db 167 DHQETLGDLSLEIAEQKAGIFKAGKKA 193

RESULT 7
US-08-764-100-25
; Sequence 25, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.
; TITLE OF INVENTION: Improvements in or Relating to Organic
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sandoz Agro, Inc
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
RESULT 2
US-08-960-022-14
; Sequence 14, Application US/08960022
; Patent No. 5976837
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Metberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,022
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 514 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: protein
; MOLECULE TYPE: protein
US-08-960-022-14

Query Match 34.1%; Score 44; DB 2; Length 514;
Best Local Similarity 36.0%; Pred. No. 67;
Matches 9; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 3 ALEALADQTDALQSEEAHVKNDA 27
Db 240 SIKAQRAQTOLLKEEARLVESNNA 264

RESULT 3
US-08-934-846-4
; Sequence 4, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folc
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
```

```
ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,846
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-934-846-4

Query Match 33.3%; Score 43; DB 2; Length 407;
Best Local Similarity 33.3%; Pred. No. 72;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 2 DALEALADQTDALQSEEAHVKNDA 28
Db 167 DHOETLGDSLEATAEQKAGIFKAGKA 193

RESULT 4
US-09-238-557-4
; Sequence 4, Application US/09238557
; Patent No. 6165472
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folc
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,557
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 02:06:51 ; Search time 4.34225 Seconds  
(without alignments)  
157.503 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALEALADQTDALQSEAAVVKADNA 28

Scoring table: BLOSUM62  
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A-COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B-COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A-COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B-COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS-COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	44	34.1	344	4	US-08-961-083-192
2	44	34.1	514	2	US-08-960-022-14
3	43	33.3	407	2	US-08-934-846-4
4	43	33.3	407	4	US-09-238-557-4
5	43	33.3	418	2	US-08-934-846-2
6	43	33.3	418	4	US-08-238-557-2
7	43	33.3	744	1	US-08-764-100-25
8	42	32.6	102	2	US-08-710-749-8
9	42	32.6	427	4	US-09-521-668B-18
10	42	32.6	427	4	US-09-521-668B-20
11	42	32.6	2890	4	US-09-413-814-67
12	42	32.6	3798	3	US-09-335-409-6
13	42	32.6	3798	4	US-09-568-102-6
14	42	32.6	3798	4	US-09-567-969-6
15	42	32.6	3798	4	US-09-568-480-6
16	42	32.6	3798	4	US-09-568-486-6
17	42	32.6	3798	4	US-09-568-472-6
18	41	31.8	370	5	PCT-US93-01676A-4
19	41	31.8	395	1	US-07-876-280-9
20	41	31.8	395	1	US-08-049-783-6
21	41	31.8	395	1	US-08-316-301A-10
22	41	31.8	395	2	US-08-904-278-10
23	41	31.8	395	4	US-09-076-137-10
24	41	31.8	395	4	US-09-222-594-10
25	41	31.8	395	5	PCT-US92-03624-10
26	40	31.0	119	4	US-08-858-207A-394
27	40	31.0	287	1	US-08-457-245-3

28 40 31.0 549 4 US-09-245-041-9 Sequence 9, Appl  
29 40 31.0 900 2 US-08-630-822A-62 Sequence 62, Appl  
30 40 31.0 900 2 US-09-005-069-62 Sequence 62, Appl  
31 40 31.0 1094 4 US-09-268-347-32 Sequence 32, Appl  
32 40 31.0 1253 1 US-08-252-966B-12 Sequence 12, Appl  
33 40 31.0 1253 1 US-07-920-281C-3 Sequence 3, Appl  
34 40 31.0 1253 4 US-08-466-277-3 Sequence 3, Appl  
35 40 31.0 1260 4 US-09-245-041-2 Sequence 2, Appl  
36 40 31.0 1261 1 US-08-252-966B-18 Sequence 18, Appl  
37 39 30.2 25 6 5169933-42 Patent No. 5169933  
38 39 30.2 58 5 PCT-US94-06655-12 Sequence 12, Appl  
39 39 30.2 288 2 US-08-961-858-6 Sequence 6, Appl  
40 39 30.2 288 3 US-09-089-593-6 Sequence 6, Appl  
41 39 30.2 288 3 US-08-950-925-4 Sequence 4, Appl  
42 39 30.2 326 1 US-07-603-133B-25 Sequence 25, Appl  
43 39 30.2 528 4 US-08-928-213B-8 Sequence 8, Appl  
44 39 30.2 770 1 US-08-445-135-2 Sequence 2, Appl  
45 39 30.2 912 5 PCT-US95-03747-2 Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-961-083-192  
; Sequence 192, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961.083  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 192:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-961-083-192

Query Match 34.1%; Score 44; DB 4; Length 344;  
Best Local Similarity 55.0%; Pred. No. 42;  
Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 8 ADOTDALQSEAAVVKADNA 27  
DB 3 ACQTDSQIERAAVSGGKA 22



**THIS PAGE BLANK (USPTO)**

Search completed: October 13, 2002, 02:10:49  
Job time : 74.5722 secs

```
OY 26 EKLALRNEERAIDELKQAIEDKEATTAIEAASSDALEALADQTDALQSEEAHVVDNA 85
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
DB 227 EQARIRECARLAQARQAAQAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 286
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
OY 86 ASDALEALADQTDALQSEEAHVVDNAASDAWEKAATPIALDVKKTKDTTPVVKKEERQ 145
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
DB 287 ESKAPESTTTTEESTSTESSTTTTEESTSTESSTTTTEESTSTESSTTTTEESTSTESST 346
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
OY 146 NVNTLPTTGEESN 158
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
DB 347 STSTESSTTTTES 359
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :

RESULT 34
O9KWR3 PRELIMINARY; PRT; 2178 AA.
ID O9KWR3
AC O9KWR3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE STREPTOCOCCAL HEMAGGLUTININ.
GN HSA.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN RP
RP SEQUENCE FROM N.A.
RC STRAIN-DL1;
RA Takahashi Y., Konishi K., Yoshikawa M.;
RT "Cloning and characterization of the gene encoding a hemagglutinin of
RT Streptococcus gordonii DL1.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB029393; BAA97453.1; -.
DR InterPro; IPR004089; Chemotaxis transducer.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
SQ SEQUENCE 2178 AA; 203507 MW; 75CC2E72F41DA8C CRC64;

Query Match 10.0%; Score 78; DB 2; Length 2178;
Best Local Similarity 21.4%; Pred. No. 22;
Matches 24; Conservative 27; Mismatches 61; Indels 0; Gaps 0;

OY 47 DKEATTAIEAASSDALEALADQTDALQSEEAHVVDNAASDALEALADQTDALQSEEAHV 106
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
DB 1471 ESASTSASVSASTSASTSASVSASTSASVSASTSASVSASTSASVSASTSASVSAS 1530
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
OY 107 VVQSDNAASDAWEKAATPIALDVKKTKDTTPVVKKEERQNTLPTTGEESN 158
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
DB 1531 TSASTSASVSASTSASTSASVSASTSASTSASVSASTSASTSASVSASTSASTSAS 1582
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :

RESULT 35
O9U4R2 PRELIMINARY; PRT; 310 AA.
ID O9U4R2
AC O9U4R2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE ANON1A3 (FRAGMENT).
GN CG1910 OR ANON1A3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRANCE 1;
RX MEDLINE=20050669; PubMed=10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
```

```
polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161729; AAD45733.1; -.
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 310 310
SQ SEQUENCE 310 AA; 32336 MW; 10CF3F4A40A1E13A CRC64;

Query Match 9.9%; Score 77; DB 5; Length 310;
Best Local Similarity 22.2%; Pred. No. 3.4;
Matches 24; Conservative 19; Mismatches 65; Indels 0; Gaps 0;

OY 24 APEKALRNEERAIDELKQAIEDKEATTAIEAASSDALEALADQTDALQSEEAHVVDNA 83
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
DB 148 APETEAAAKTTESNDVVVEATGKATVLEVPAAEPKAEKAEKAEKAEKAEKAEKAEKAE 207
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
OY 84 NAASDALEALADQTDALQSEEAHVVDNAASDAWEKAATPIALDVKK 131
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
DB 208 KEAASCEEPKSVVDVSAEPKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 255
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :

RESULT 36
O9I255 PRELIMINARY; PRT; 1110 AA.
ID O9I255
AC O9I255;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE NF-180.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-CNS;
RX MEDLINE=95287814; PubMed=7770000;
RA Jacobs A.J., Kamholz J., Selzer M.E.;
RT "The single lamprey neurofilament subunit (NF-180) lacks
RT multiphosphorylation repeats and is expressed selectively in
RT projection neurons.";
RL Brain Res. Mol. Brain Res. 29:43-52(1995).
DR EMBL; U19361; AAA80106.1; -.
DR InterPro; IPR001664; IF.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR PROSITE; PS00226; IF; UNKNOWN.1.
SQ SEQUENCE 1110 AA; 123818 MW; 6558DA73DAF6974C CRC64;

Query Match 9.9%; Score 77; DB 13; Length 1110;
Best Local Similarity 23.1%; Pred. No. 13;
Matches 21; Conservative 24; Mismatches 46; Indels 0; Gaps 0;

OY 55 EAASDALEALADQTDALQSEEAHVVDNAASDALEALADQTDALQSEEAHVVDNA 114
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
DB 706 EAEEVEEVEEVTKSDBEAEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 765
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
OY 115 SDANEKAATPIALDVKKTKDTTPVVKKEERQ 145
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
DB 766 EEAKEASDDEKPEEEVKESEAPVAPPAKK 796
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :

RESULT 37
O21022 PRELIMINARY; PRT; 1133 AA.
ID O21022
AC O21022; Q21071;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE F59A2.6 PROTEIN.
GN F59A2.6.
OS Caenorhabditis elegans.
```



```

RESULT 27
Q93EM8 ID Q93EM8 PRELIMINARY; PRT; 669 AA.
AC Q93EM8;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DE MIG PRECURSOR.
OS Streptococcus dysgalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43078;
RX MEDLINE=21437603; PubMed=11553540;
RA Song X.M., Perez-Casal J., Bolton A., Potter A.A.;
RT "Surface-Expressed Mlg Protein Protects Streptococcus dysgalactiae
RT against Phagocytosis by Bovine Neutrophils.";
RL Infect. Immun. 69:6030-6037(2001).
DR EMBL: AF354651; AAL09476.1;
KW Signal.
FT SIGNAL.
FT CHAIN 1 39 POTENTIAL.
FT CHAIN 40 669 MIG.
SQ SEQUENCE 669 AA; 72682 MW; 5C8982B952029B03 CRC64;

Query Match 10.4%; Score 81; DB 2; Length 669;
Best Local Similarity 53.1%; Pred. No. 2.9;
Matches 17; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 128 DVKTKTKTPVVKKEQRQNVNTLPTTGESNP 159
: | | | | | | | | | | | | | | | | | |
Db 613 EAKPEAKKEEAKKEAKKATLPTTGESNP 644

RESULT 28
Q9F553 ID Q9F553 PRELIMINARY; PRT; 1272 AA.
AC Q9F553;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE RELAXASE.
GN NIKB.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K-12; TRANSPOSON-TN7;
RA Sempel G., Motomura K., Masuda S., Yamaguchi T., Ando K., Oishi T.,
RA Furuya N., Komano T., Mizobuchi K.;
RT "Organization and diversification of plasmid genomes: complete
RT nucleotide sequence of the R721 genome.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K-12; TRANSPOSON-TN7;
RX MEDLINE=93015772; PubMed=1400257;
RA Kim S., Komano T.;
RT "Nucleotide sequence of the R721 shuffle.";
RL J. Bacteriol. 174:7053-7058(1992).
DR EMBL: AP002527; BAB12624.1;
KW Plasmid.
SQ SEQUENCE 1272 AA; 145804 MW; 30A4839CA2519BA7 CRC64;

Query Match 10.4%; Score 81; DB 2; Length 1272;
Best Local Similarity 27.3%; Pred. No. 5.9;
Matches 21; Conservative 13; Mismatches 43; Indels 0; Gaps 0;

```

```

QY 38 DELKQKQIEDKEATTATIEAASSDALEALADQTDALQSEEAAYVVKADNAASDALEALADQT 97
: | | | | | | | | | | | | | | | | | |
Db 806 DALKEAQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ 865

QY 98 DALQSEEAAYVQSDNAA 114
: | | | | | | | | | | | | | | | | | |
Db 866 DTEQAQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ 865

RESULT 29
Q9CQ47 ID Q9CQ47 PRELIMINARY; PRT; 274 AA.
AC Q9CQ47;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE 4921530L21RIK PROTEIN.
GN 4921530L21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Harai A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Sakimi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fietcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN [1]
RP Nature 409:685-690(2001).
DR EMBL: AK014987; BAB29656.1;
DR EMBL: AK005846; BAB24271.1;
DR MGD; MGI:1913982; 4921530L21RIK.
SQ SEQUENCE 274 AA; 32075 MW; C810F3986D46B5C2 CRC64;

Query Match 10.3%; Score 80; DB 11; Length 274;
Best Local Similarity 29.2%; Pred. No. 1.4;
Matches 21; Conservative 18; Mismatches 33; Indels 0; Gaps 0;

QY 33 EERAIDELKKQAIEDKEATTATIEAASSDALEALADQTDALQSEEAAYVVKADNAASDALEA 92
: | | | | | | | | | | | | | | | | | |
Db 162 EETKALQHOERALRWEERKALMRDGVAAELAEALTKEGAALMEEQALWKKEQALREENKA 221

QY 93 LADQTDALQSEE 104
: | | | | | | | | | | | | | | | | | |
Db 222 LREHGALQDEE 233

RESULT 30
Q9WWX1 ID Q9WWX1 PRELIMINARY; PRT; 372 AA.
AC Q9WWX1;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE TOLA PROTEIN.

```

[illegible]

```

RP SEQUENCE FROM N.A.
RX STRAIN=PADUA/ITALY 15, AND PADUA/ITALY 14;
RX MEDLINE=20050669; PubMed=10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161738; AAD45743.1; -
DR EMBL; AF161738; AAD45742.1; -
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 1
FT NON_TER 310
SQ SEQUENCE 310 AA; 32392 MW; F7657B8CDA2F81E CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

OY 24 APEKLALEADOTDALQSEAEVQSDNAADAWKATPIALDYK 83
DB 148 APEAAKTTESNDVVEATGKTVLEVPAAEPKEASTVESAEELTETSTVVVTEP 207

OY 84 NAASDALEALADOTDALQSEAEVQSDNAADAWKATPIALDYK 131
DB 208 KEAASSEEPSKVVDSPAEPKESNTDESATPPIDVSTAPASNDVSK 255

RESULT 22
OY9VU3
ID Q9VU3 PRELIMINARY; PRT; 489 AA.
AC Q9VU3; Q9TW86;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE ANON-EST:FEI33 PROTEIN (ANONIA3).
GN CG1910 OR ANON-EST:FEI33 OR ANONIA3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Anil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berhan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

```

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE OF 180-489 FROM N.A.
RX STRAIN-VARIOUS STRAINS;
RX MEDLINE=20050669; PubMed=10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF003779; AAF57190.1; -
DR EMBL; AF161747; AAD45751.1; -
DR EMBL; AF161723; AAD45727.1; -
DR EMBL; AF161727; AAD45731.1; -
DR EMBL; AF161732; AAD45736.1; -
DR EMBL; AF161737; AAD45741.1; -
DR EMBL; AF161740; AAD45744.1; -
DR EMBL; AF161741; AAD45745.1; -
DR EMBL; AF161742; AAD45746.1; -
DR EMBL; AF161745; AAD45749.1; -
DR EMBL; AF161746; AAD45750.1; -
DR FlyBase; FBgn0022349; CG1910.
SQ SEQUENCE 489 AA; 51587 MW; 85BBFA05060A9FD2 CRC64;

Query Match 10.8%; Score 84; DB 5; Length 489;
Best Local Similarity 23.1%; Pred. No. 1;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

OY 24 APEKLALEADOTDALQSEAEVQSDNAADAWKATPIALDYK 83
DB 327 APEAAKTTESNDVVEATGKTVLEVPAAEPKEASTVESAEELTETSTVVVTEP 386

OY 84 NAASDALEALADOTDALQSEAEVQSDNAADAWKATPIALDYK 131
DB 387 KEAASSEEPSKVVDSPAEPKESNTDESATPPIDVSTAPASNDVSK 434

RESULT 23
OY9G0H8
ID Q9G0H8 PRELIMINARY; PRT; 540 AA.
AC Q9G0H8;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE GP7.
OS Roseophage SIOL.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like phages.
OX NCBI_TaxID=136084;
RN [1]
RP SEQUENCE FROM N.A.
RA Rohwer F.L., Segall A.M., Steward G., Seguritan V., Breitbart M.,
RA Wolven F., Azam F.;
RT "The complete genomic sequence of the marine phage Roseophage SIOL
RL Limnol. Oceanogr. 45:408-418(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Rohwer F.L., Segall A.M., Steward G., Seguritan V., Breitbart M.,
RA Wolven F., Azam F.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF189021; AAG02590.1; -
DR InterPro; IPR000104; Antifreeze_1.

```





```

RX MEDLINE-20050669; PubMed-10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161724; AAD45728.1; -.
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 1
FT NON_TER 310
SQ SEQUENCE 310 AA; 32391 MW; 00EBE2551606CB57 CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

QY 24 APEKALRNEERAIDELKKOAIKEDKEATTAIEAASSDALEALADOTDALQSEEAAYVKAD 83
Db 148 APETEAANKTSSNDVVEVATEGKTVLEVPAAEPKEAESTVESAEELTETSTVVVTEP 207

QY 84 NAASDALEALADOTDALQSEEAAYVQSDNAASDAWEKAATPIALDYKK 131
Db 208 KEAASSEEPKSVVDSEAPKEAESNTDESATVPIDVSTAPASNDYSK 255

RESULT 15
Q9U4R4
ID Q9U4R4 PRELIMINARY; PRT; 310 AA.
AC Q9U4R4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ANONIA3 (FRAGMENT).
GN CG1910 OR ANONIA3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AUSTRALIA 5;
RX MEDLINE-20050669; PubMed-10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161726; AAD45730.1; -.
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 1
FT NON_TER 310
SQ SEQUENCE 310 AA; 32372 MW; 122BE25517DABA90 CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

QY 24 APEKALRNEERAIDELKKOAIKEDKEATTAIEAASSDALEALADOTDALQSEEAAYVKAD 83
Db 148 APETEAANKTSSNDVVEVATEGKTVLEVPAAEPKEAESTVESAEELTETSTVVVTEP 207

QY 84 NAASDALEALADOTDALQSEEAAYVQSDNAASDAWEKAATPIALDYKK 131
Db 208 KEAASSEEPKSVVDSEAPKEAESNTDESATVPIDVSTAPASNDYSK 255

RESULT 16
Q9U4R3
ID Q9U4R3 PRELIMINARY; PRT; 310 AA.
AC Q9U4R3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

```

```

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ANONIA3 (FRAGMENT).
GN CG1910 OR ANONIA3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CYPRIUS;
RX MEDLINE-20050669; PubMed-10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161728; AAD45732.1; -.
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 1
FT NON_TER 310
SQ SEQUENCE 310 AA; 32364 MW; 0CCC02A07AA114A1 CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

QY 24 APEKALRNEERAIDELKKOAIKEDKEATTAIEAASSDALEALADOTDALQSEEAAYVKAD 83
Db 148 APETEAANKTSSNDVVEVATEGKTVLEVPAAEPKEAESTVESAEELTETSTVVVTEP 207

QY 84 NAASDALEALADOTDALQSEEAAYVQSDNAASDAWEKAATPIALDYKK 131
Db 208 KEAASSEEPKSVVDSEAPKEAESNTDESATVPIDVSTAPASNDYSK 255

RESULT 17
Q9U4R0
ID Q9U4R0 PRELIMINARY; PRT; 310 AA.
AC Q9U4R0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ANONIA3 (FRAGMENT).
GN CG1910 OR ANONIA3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ITALY;
RX MEDLINE-20050669; PubMed-10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161731; AAD45735.1; -.
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 1
FT NON_TER 310
SQ SEQUENCE 310 AA; 32306 MW; 0EA033FEDA5D25FE CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

QY 24 APEKALRNEERAIDELKKOAIKEDKEATTAIEAASSDALEALADOTDALQSEEAAYVKAD 83
Db 148 APETEAANKTSSNDVVEVATEGKTVLEVPAAEPKEAESTVESAEELTETSTVVVTEP 207

```



RA	Baker S., Basham D., Brooks K., Chillingworth T., Connerton P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT	*Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.*;
RL	Nature 413:848-852(2001).
RN	EMBL; AF513384; CAD09918.1; -.
KW	Plasmid.
SQ	SEQUENCE 850 AA; 90917 MW; 21271493A07A6563 CRC64;
 Query Match 13.4%; Score 104; DB 2; Length 850; Best Local Similarity 34.6%; Pred. No. 0.014; Matches 28; Conservative 12; Mismatches 41; Indels 0; Gaps	
QY	49 EATTAETAEASSDALEALADOTDALOSEEAUVKADNAASDALEALADOTDALOSEEAUVV 108    :     :   :   :    :    :    :    :    :    :    :
Db	169 EAUNSKQAAGSEANAKSANAASQAARATESNARASKERAAASQTAALOSEQOAAA 228 ::  :
QY	109 QSDNAASDAWEKAATPIADV 129 :
Db	229 HADSASEAEKAKKFADLDV 249 :
 RESULT 9 Q53974 PRELIMINARY; PRT; 413 AA. ID Q53974 AC Q53974 Q53974; DT 01-NOV-1996 (TrEMBLrel. 01, Created) DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update) DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update) DE MAG PROTEIN PRECURSOR. GN MAG. OS Streptococcus dysgalactiae. OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; OX Streptococcus. NCBI_TaxID=1334; [1] RP SEQUENCE FROM N.A. RX MEDLINE=94259307; PubMed=7515368; RT Jonsson H., Frykberg L., Rantamaeki L., Guss B.; RT *MAG, a novel plasma protein receptor from Streptococcus RL dysgalactiae.*; RL Gene 143:85-89(1994). DR EMBL; L27798; AAA26921.1; -. DR HSSP; P06654; IPGX. DR InterPro; IPR002988; GA. DR InterPro; IPR001899; Gram_pos_anchor. DR InterPro; IPR000724; IgG_bind_B. DR Pfam; PF01468; GA; 1. DR Pfam; PF00746; Gram_pos_anchor; 1. DR Pfam; PF01378; IgG_binding_B; 1. DR PRINTS; PS00015; GPOSANCHOR. DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1. KW Signal; Transmembrane. FT SIGNAL 1 34 POTENTIAL. FT CHAIN 35 413 POTENTIAL. SQ SEQUENCE 413 AA; 44004 MW; F04DC7104AF9E50F CRC64;	
 Query Match 11.7%; Score 91; DB 2; Length 413; Best Local Similarity 59.4%; Pred. No. 0.15; Matches 19; Conservative 1; Mismatches 12; Indels 0; Gaps	
QY	128 DVKTKDTKTPVKKEQRQNVTTLPTTGEESNP 159 
Db	357 DTRKEDARKPEAKKEAKAATLTPTTGEESNP 388 
 RESULT 10 Q95S18 PRELIMINARY; PRT; 619 AA. ID Q95S18	

```
Oy 121 AA 122
Db 121 LA 122

RESULT 5
Q9X5C5
ID Q9X5C5 PRELIMINARY; PRT; 239 AA.
AC Q9X5C5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KTL9;
RX MEDLINE=95269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT *Protein GRAB of streptococcus pyogenes regulates proteolysis at the
RT bacterial surface by binding alpha2-macroglobulin.*;
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL: AF124402; AAD26341.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 239
SQ SEQUENCE 239 AA; 24846 MW; 61AC4F6F863AF0F5 CRC64;

Query Match 68.7%; Score 534; DB 2; Length 239;
Best Local Similarity 91.8%; Pred. No. 6.2e-49;
Matches 112; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Oy 1 VDSPIEQRIIPNGGTLNLLGNAPKALNRNEAIDELKKQATDEKATTAIEAASD 60
Db 1 VDSPIEQRIIPNGGTLNLLGNAPKALNRNEAIDELKKQATDEKATTAIEAASD 60

Oy 61 ALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVQSDNAASDAWEK 120
Db 61 ALEALADQADALQSEAAVQSDNAASDALEALADQTDALQSEAAVVKADNAASDTEA 120

Oy 121 AA 122
Db 121 LA 122

RESULT 6
Q54180
ID Q54180 PRELIMINARY; PRT; 103 AA.
AC Q54180;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PROTEIN G'.
GN PROTEIN G' GENE.
OS Streptococcus sp. 'group G'.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G148;
RX MEDLINE=90226312; PubMed=2183792;
RA Goward C.R., Murphy J.P., Atkinson T., Barstow D.A.;
RT *Expression and purification of a truncated recombinant streptococcal
RT protein G.*;
RL Biochem. J. 267:171-177(1990).
DR EMBL: X53324; CAA37409.1; -.
SQ SEQUENCE 103 AA; 10828 MW; E7AE8BD1C073423C CRC64;
```

```
Query Match 18.3%; Score 142; DB 2; Length 103;
Best Local Similarity 57.9%; Pred. No. 1.3e-07;
Matches 33; Conservative 3; Mismatches 21; Indels 0; Gaps 0;

Oy 1 VDSPIEQRIIPNGGTLNLLGNAPKALNRNEAIDELKKQATDEKATTAIEA 57
Db 34 VDSPIEDPIIRNGGELTNLLGNSETTLALNRERSATADLTAAAVADTVAAAAAENA 90

RESULT 7
Q53900
ID Q53900 PRELIMINARY; PRT; 323 AA.
AC Q53900;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ALBUMIN-BINDING PROTEIN (FRAGMENT).
OS Streptococcus canis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DG12;
RX MEDLINE=92363555; PubMed=1500168;
RA Sjoberg U.;
RT *Isolation and molecular characterization of a novel albumin-binding
RT protein from group G streptococci.*;
RL Infect. Immun. 60:3601-3608(1992).
DR EMBL: M95520; AAA26847.1; -.
DR HSSP: Q51911; IGAB.
DR InterPro: IPR002988; GA.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF01469; GA; 2.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 323
SQ SEQUENCE 323 AA; 34990 MW; E6903CFDCD5D373E CRC64;

Query Match 16.2%; Score 126; DB 2; Length 323;
Best Local Similarity 42.6%; Pred. No. 2.2e-05;
Matches 26; Conservative 14; Mismatches 21; Indels 0; Gaps 0;

Oy 62 LEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVQSDNAASDAWEKA 121
Db 62 LSAQADQIVSAQADNEAITKAEDESSKAWAAADQANTAKAEADELAKAEKSSDAWEKA 121

Oy 122 A 122
Db 122 A 122

RESULT 8
Q935B1
ID Q935B1 PRELIMINARY; PRT; 850 AA.
AC Q935B1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PUTATIVE PHAGE TAIL PROTEIN.
GN HCM2.0051C.
OS Salmonella enterica subsp. enterica serovar Typhi.
OX Plasmid pHC2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=90370;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
```

```
FT CHAIN 34 217 GRAB.
SQ SEQUENCE 217 AA; 22836 MW; 79AA8C4FF5F3FA06 CRC64;

Query Match 100.0%; Score 777; DB 16; Length 217;
Best Local Similarity 100.0%; Pred. No. 7.9e-75;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKAIEDKEATTAEAASSD 60
Db 34 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKAIEDKEATTAEAASSD 93
Qy 61 ALEALADQTDALQSEEAHVKAADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 94 ALEALADQTDALQSEEAHVKAADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 153
Qy 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEENP 159
Db 154 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEESNP 192

RESULT 2
Q9S6G4 PRELIMINARY; PRT; 156 AA.
AC Q9S6G4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KTL3;
RX MEDLINE=99269061; PubMed=10336419;
RT *Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.*;
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL: AF124401; AAD26340.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 156
SQ SEQUENCE 156 AA; 16575 MW; 0C20967E2FB0D866 CRC64;

Query Match 97.8%; Score 760; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.6e-73;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKAIEDKEATTAEAASSD 60
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKAIEDKEATTAEAASSD 60
Qy 61 ALEALADQTDALQSEEAHVKAADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 61 ALEALADQTDALQSEEAHVKAADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Qy 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEENP 156
Db 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEENP 156

RESULT 3
Q9S6G3 PRELIMINARY; PRT; 156 AA.
AC Q9S6G3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
```

```
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AP1;
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT *Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.*;
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL: AF124400; AAD26339.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 156
SQ SEQUENCE 156 AA; 16587 MW; 19FAA98D0599D866 CRC64;

Query Match 97.0%; Score 754; DB 2; Length 156;
Best Local Similarity 99.4%; Pred. No. 1.5e-72;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKAIEDKEATTAEAASSD 60
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKAIEDKEATTAEAASSD 60
Qy 61 ALEALADQTDALQSEEAHVKAADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 61 ALEALADQTDALQSEEAHVKAADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Qy 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEENP 156
Db 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEENP 156

RESULT 4
Q9X5C6 PRELIMINARY; PRT; 268 AA.
AC Q9X5C6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AP49;
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT *Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.*;
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL: AF124403; AAD26342.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 268
SQ SEQUENCE 268 AA; 27744 MW; 1C59239260CD7E7 CRC64;

Query Match 69.0%; Score 536; DB 2; Length 268;
Best Local Similarity 92.6%; Pred. No. 4.3e-49;
Matches 113; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKAIEDKEATTAEAASSD 60
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKAIEDKEATTAEAASSD 60
Qy 61 ALEALADQTDALQSEEAHVKAADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 61 ALEALADQTDALQSEEAHVKAADNAASDALEALADQTDALQSEEAHVQSDNAAGDALEA 120
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 01:14:06 ; Search time 70.5722 Seconds  
(without alignments)  
389.760 Million cell updates/sec

Title: US-09-847-539A-6  
Perfect score: 777  
Sequence: 1 VDSPIEQRIIPNGTITNL.....KKEERQVNTLPTTGESNP 159

Scoring table: BLOSUM62

Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacterioph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	777	100.0	217	16 Q9X5C4	Q9X5C4 streptococc
2	760	97.8	156	2 Q9S6G4	Q9S6G4 streptococc
3	754	97.0	156	2 Q9S6G3	Q9S6G3 streptococc
4	536	69.0	268	2 Q9X5C6	Q9X5C6 streptococc
5	534	68.7	239	2 Q9X5C5	Q9X5C5 streptococc
6	142	18.3	103	2 Q94180	Q94180 streptococc
7	126	16.2	323	2 Q93900	Q93900 streptococc
8	104	13.4	850	2 Q935B1	Q935B1 salmonella
9	91	11.7	413	2 Q93974	Q93974 streptococc
10	90	11.6	619	5 Q9S18	Q9S18 drosophila
11	90	11.6	2148	5 Q9BJD3	Q9BJD3 physarum po
12	89	11.5	447	2 Q9AKA4	Q9AKA4 streptococc
13	85	10.9	664	2 Q93975	Q93975 streptococc
14	84	10.8	310	5 Q9U4R5	Q9U4R5 drosophila
15	84	10.8	310	5 Q9U4R4	Q9U4R4 drosophila
16	84	10.8	310	5 Q9U4R3	Q9U4R3 drosophila

17 84 10.8 310 5 Q9U4R0  
18 84 10.8 310 5 Q9U4Q9  
19 84 10.8 310 5 Q9U4Q8  
20 84 10.8 310 5 Q9U4Q7  
21 84 10.8 310 5 Q9TVG2  
22 84 10.8 489 5 Q9V9U3  
23 84 10.8 540 9 Q9G0H8  
24 83 10.7 310 5 Q9TW87  
25 81 10.4 310 5 Q9U4Q6  
26 81 10.4 558 2 Q94718  
27 81 10.4 669 2 Q93EM8  
28 81 10.4 1272 2 Q9E553  
29 80 10.3 274 11 Q9CQ47  
30 80 10.3 372 2 Q9WXX1  
31 80 10.3 3072 2 Q939N5  
32 78 10.0 113 3 Q98245  
33 78 10.0 576 2 Q9KJ33  
34 78 10.0 2178 2 Q9KWR3  
35 77 9.9 310 5 Q9U4R2  
36 77 9.9 1110 13 Q91255  
37 77 9.9 1133 5 Q21022  
38 76 9.8 395 2 Q937K4  
39 76 9.8 725 13 Q07718  
40 76 9.8 2314 12 Q91LF6  
41 75 9.7 211 11 Q99JX1  
42 75 9.7 509 10 Q9C8J7  
43 75 9.7 791 13 Q9DGL1  
44 75 9.7 969 5 Q9NDI9  
45 75 9.7 2478 2 Q9LCH2

#### ALIGNMENTS

RESULT 1

Q9X5C4 PRELIMINARY; PRT; 217 AA.  
AC Q9X5C4;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE GRAB PRECURSOR (PROTEIN GRAB) (PROTEIN G-RELATED ALPHA 2N-BINDING PROTEIN).  
DE GRAB OR SPY1357.  
GS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RN [2]  
RP STRAIN-ATCC700294;  
RA Rasmussen M., Muller H.P., Bjorck L.;  
RT "Protein GRAB of Streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
DR EMBL; AF124399; AAD26338.1;  
DR EMBL; AE006573; AAK34185.1;  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR PRINTS; PS00015; GPOSANCHOR.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Complete proteome; Signal; Transmembrane.  
FT SIGNAL 1 33  
POTENTIAL.

THIS PAGE BLANK (USPTO)

APPLICANT: SHARMA, Lawrence Rajendra  
TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR  
TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR  
TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THEREOF, AND USES THEREOF  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,847  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/127,499  
FILING DATE: 28-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 51916/104/INBI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 695 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
US-08-482-847-23

Query Match 11.6%; Score 90; DB 1; Length 695;  
Best Local Similarity 24.2%; Pred. No. 0.89;  
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;  
QY 24 APEKALRNEERA-IDELKKQATEDKEATTAIEAASSDALEALADQTDALQSEAAV--- 79  
Db 212 APELTKKLEAKLEAEAKKATEAKOKVDAAEVAPQAKIAELENGVHRLEQELKEIDES 271  
QY 80 -----VKADNAASDALEALADQTDALQSEAAEVVQSDNAA----- 114  
Db 272 ESEDYAKEGFRAPLQSKLDKAKKLSKLELSDKIDELDAEIAKLEDLKAAEENNVED 331  
QY 115 --SDAWEKAATPIALDVKKTK-DTKPVVKKKEER 144  
Db 332 YFKEGLEKTIAAKKAELKTEADLKKAVNEPEK 364

Search completed: October 13, 2002, 04:50:06  
Job time : 28.6378 secs



```
/ APPLICATION NUMBER: US/08/247,491A
/ FILING DATE: 23-JUN-1994
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: KOWALSKI, Thomas J.
/ REGISTRATION NUMBER: 32,147
/ REFERENCE/DOCKET NUMBER: 454312-2041
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 212-588-0800
/ TELEFAX: 212-588-0500
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 648 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-247-491A-2

Query Match 11.6%; Score 90; DB 2; Length 648;
Best Local Similarity 24.2%; Pred. No. 0.81;
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;

QY 24 APEKALRNEERA-IDELKKQATIEDKATTAEAASSDALEALADQTDALQSEEA--- 79
Db 170 APELTKKLEAKLEAEKATEAKQKVDAEVAPOAKIAELENOVHRLQELKEIDES 229

QY 80 -----VKADNAASDALEALADQTDALQSEEA--- 114
Db 230 ESDYAKEGFRAPLOSKLDKAKKLSKLELSKIDELDAETAKLEDQLKAAEENNVED 289

QY 115 --SDAWEKAATPIALDVKKTK-DTRKPVVKKER 144
Db 290 YFKEGLEKTIAAKKALEKTEADLKKAVNEPEK 322

RESULT 38
US-08-127-499A-23
/ Sequence 23; Application US/08127499A
/ Patent No. 5510284
/ GENERAL INFORMATION:
/ APPLICANT: VAN ALSTYNE, Diane
/ TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED
/ TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES
/ NUMBER OF SEQUENCES: 40
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Foley & Lardner
/ STREET: 3000 K Street, N.W., Suite 500
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: USA
/ ZIP: 20007-5109
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ FILING DATE: 28-SEP-1993
/ APPLICATION NUMBER: US/08/127,499A
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BENT, Stephen A.
/ REGISTRATION NUMBER: 29,768
/ REFERENCE/DOCKET NUMBER: 51916/102/INBI
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202)672-5300
/ TELEFAX: (202)672-5399
/ TELEX: 904136
/ INFORMATION FOR SEQ ID NO: 23:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 695 amino acids
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: unknown
/ US-08-127-499A-23

Query Match 11.6%; Score 90; DB 1; Length 695;
Best Local Similarity 24.2%; Pred. No. 0.89;
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;

QY 24 APEKALRNEERA-IDELKKQATIEDKATTAEAASSDALEALADQTDALQSEEA--- 79
Db 212 APELTKKLEAKLEAEKATEAKQKVDAEVAPOAKIAELENOVHRLQELKEIDES 271

QY 80 -----VKADNAASDALEALADQTDALQSEEA--- 114
Db 272 ESDYAKEGFRAPLOSKLDKAKKLSKLELSKIDELDAETAKLEDQLKAAEENNVED 331

QY 115 --SDAWEKAATPIALDVKKTK-DTRKPVVKKER 144
Db 332 YFKEGLEKTIAAKKALEKTEADLKKAVNEPEK 364

RESULT 40
US-08-482-847-23
/ Sequence 23; Application US/08482847
/ Patent No. 5556757
/ GENERAL INFORMATION:
/ APPLICANT: VAN ALSTYNE, Diane
```

```

; STREET: Suite 1203, 2001 Jefferson Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,718
; FILING DATE:
;
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,068
; FILING DATE: 03 JUNE 1993
; APPLICATION NUMBER: US/07/835,698
; FILING DATE: 12-FEB-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/656,773
; FILING DATE: 15-FEB-1991
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 521-0378
; TELEX: LUKPAT WASHINGTON
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 648 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-468-718-2
;
; Query Match 11.6%; Score 90; DB 2: Length 648;
; Best Local similarity 24.2%; Pred. No. 0.81;
; Matches 37; Conservative 23; Mismatches 61; Indels 3
;
QY 24 APEKALRNERR-IDELKKQATIEDKEATTATIEAASSDALEALADQTDALQSEEA
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 170 APELTKKLEEAKEAEKATKOKVDAAEVAFOAKTAELENOVHRLEQELKI
    QY 80 -----VKADNNAASDALEALADQTDALQSEEAEEVVSQDNAA-----
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230 ESEDYAKEGFRAPLOSKLDAKKAKLSKLELSDKIDELDAEIAKLEDLQAKAEEN
    QY 115 --SDAWEKAATPIALOVKTKT-DTKPPVVKKEER 144
    : || :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 290 YFKEGLEKTTIAAKKAKELEKTEADLKKAVNEPEK 322
    QY 37
;
; US-08-247-491A-2
; Sequence 2, Application US/08247491A
; Patent No. 5965400
; GENERAL INFORMATION:
; APPLICANT: BRILES, David E.
; APPLICANT: YOTHER, Janet L.
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10151
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:

```

TELEX: LUKPAT WASHINGTON  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 648 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-072-070-2

Query Match 11.6% Score 90; DB 1; Length 648;  
Best Local Similarity 24.2% Pred. No. 0.81;  
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;  
QY 24 APEKALRNEERA-IDELKKAIEDKATTAIEAASDSEALADQTDALQSEAAV--- 79  
Db 170 APELTKKLEAKALEAEKKAATEAKQVDAEEVAPQAKIAELENOVHRLQEQLKEIDES 229  
QY 80 -----VKADNAASDALEALADQTDALQSEAEVQSDNAA----- 114  
Db 230 ESEDYAKGFRAPLQSKLDKAKKLSLELSKIDELDAEIAKLEDLQKAAEENNNVED 289  
QY 115 --SDAWEKAATPIALDVKKTK-DTPKPVVKKER 144  
Db 290 YFKEGLEKTTIAAKKAELEKTEADLKAVNEPEK 322

## RESULT 33

US-08-469-434-2  
Sequence 2, Application US/08469434  
Patent No. 5753463  
GENERAL INFORMATION:  
APPLICANT: Briles, David E  
APPLICANT: Yother, Janet L  
APPLICANT: McDaniel, Larry S  
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: Suite 1203, 2001 Jefferson Davis Highway  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,434  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/072,065  
FILING DATE: 03 JUNE 1993  
APPLICATION NUMBER: US/07/835,698  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/656,773  
FILING DATE: 15-FEB-1991  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 521-0378  
TELEX: LUKPAT WASHINGTON  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 648 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-469-434-2

Query Match 11.6% Score 90; DB 1; Length 648;  
Best Local Similarity 24.2% Pred. No. 0.81;  
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;  
QY 24 APEKALRNEERA-IDELKKAIEDKATTAIEAASDSEALADQTDALQSEAAV--- 79  
Db 170 APELTKKLEAKALEAEKKAATEAKQVDAEEVAPQAKIAELENOVHRLQEQLKEIDES 229  
QY 80 -----VKADNAASDALEALADQTDALQSEAEVQSDNAA----- 114  
Db 230 ESEDYAKGFRAPLQSKLDKAKKLSLELSKIDELDAEIAKLEDLQKAAEENNNVED 289  
QY 115 --SDAWEKAATPIALDVKKTK-DTPKPVVKKER 144  
Db 290 YFKEGLEKTTIAAKKAELEKTEADLKAVNEPEK 322

## RESULT 34

US-08-214-222-2  
Sequence 2, Application US/08214222  
Patent No. 5804193  
GENERAL INFORMATION:  
APPLICANT: Briles, David E  
APPLICANT: Yother, Janet L  
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: Suite 1203, 2001 Jefferson Davis Highway  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,222  
FILING DATE: 17-MAR-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/835,698  
FILING DATE: 12-FEB-1992  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 521-0378  
TELEX: LUKPAT WASHINGTON  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 648 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-214-222-2  
Query Match 11.6% Score 90; DB 1; Length 648;  
Best Local Similarity 24.2% Pred. No. 0.81;  
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;  
QY 24 APEKALRNEERA-IDELKKAIEDKATTAIEAASDSEALADQTDALQSEAAV--- 79  
Db 170 APELTKKLEAKALEAEKKAATEAKQVDAEEVAPQAKIAELENOVHRLQEQLKEIDES 229  
QY 80 -----VKADNAASDALEALADQTDALQSEAEVQSDNAA----- 114  
Db 230 ESEDYAKGFRAPLQSKLDKAKKLSLELSKIDELDAEIAKLEDLQKAAEENNNVED 289  
QY 115 --SDAWEKAATPIALDVKKTK-DTPKPVVKKER 144  
Db 290 YFKEGLEKTTIAAKKAELEKTEADLKAVNEPEK 322





[illegible]

STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/106.981  
FILING DATE: 16-AUG-1993  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: WILSON, MARY J.  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 163-23  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
TELEX: 200797 NIXN UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190  
TYPE: Amino acid  
TOPOLOGY: Unknown  
MOLECULE TYPE: Protein  
ORIGINAL SOURCE: Triticum Aestivum L.  
US-08-106-981-2

Query Match 11.6%; Score 90; DB 1; Length 190;  
Best Local Similarity 31.0%; Pred. No. 0.16;  
Matches 36; Conservative 15; Mismatches 45; Indels 20; Gaps 5;

QY 7 PRIPNGGTNLNIGNAPEKALRNE-----ERAI-----DELKKAQTEDEKAT-- 52  
DB 77 RPRTV-SGRL--CLQNPATPAYNDADATDKAIDGVGVADLKGVAEAEAVSGN 133  
QY 53 ---ATEAASSDALEALADQTDALQSEEAHVKNADNASDALEALADQTDALQSEEA 105  
DB 134 TEKAAEEAGKASEVDKAKDFGEQAKKATEEAWDGAQGIIDKVAAGAAKREA 189

RESULT 24  
US-08-465-746-2  
Sequence 2, Application US/08465746  
Patent No. 5679768  
GENERAL INFORMATION:  
APPLICANT: Briles, David E  
APPLICANT: Yother, Janet L  
APPLICANT: McDaniel, Larry S  
TITLE OF INVENTION: EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE  
TITLE OF INVENTION: PROTEIN A  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: Suite 1203, 2001 Jefferson Davis Highway  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465.746  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/048.896  
FILING DATE:  
APPLICATION NUMBER: US 07/656.773

FILING DATE: 15-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/835.698  
FILING DATE: 12-FEB-1992  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 521-0378  
TELEX: LUKPAT WASHINGTON  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 619 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-465-746-2  
Query Match 11.6%; Score 90; DB 1; Length 619;  
Best Local Similarity 24.2%; Pred. No. 0.76;  
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;  
QY 24 APEKLALRNEERA-IDELKKAQIEDKEATTATAEASSDALEALADQTDALQSEEAHV--- 79  
DB 170 APELTKKLEEAKEAEAEKATEAKQKVDAEEVAPQAKIAELENOVHRLBOELKEIDES 229  
QY 80 -----VKADNAAASDALEALADQTDALQSEEAHVQSDNAA----- 114  
DB 230 ESEDYAKEGFRAPLQSKLDKAKKLSLELSDKIDELDAETAKLELDQKAAEENNVED 289  
QY 115 --SDAWEKAATPIALDVKKTK-DTKPVVKKBER 144  
DB 290 YFKEGLEKTIAAKRAELEKTEADLUKAVNEPEK 322  
RESULT 25  
US-08-214-164-2  
Sequence 2, Application US/08214164  
Patent No. 5728387  
GENERAL INFORMATION:  
APPLICANT: BRILES, DAVID E.  
APPLICANT: YOTHER, JANET L.  
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: Suite 1203, 2001 Jefferson Davis Highway  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/214.164  
FILING DATE: 17-MAR-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/656.773  
FILING DATE: 15-FEB-1991  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W.  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 6102-137  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 521-0813  
TELEX: LUKPAT WASHINGTON  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 619 amino acids









; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 472 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-115-746-10

Query Match 11.9%; Score 92.5; DB 4; Length 472;  
Best Local Similarity 22.2%; Pred. No. 0.3; Mismatches 36; Indels 11; Gaps 4;  
Matches 34; Conservative 36; Mismatches 36; Indels 11; Gaps 4;

Qy 3 SPIEQRIIPNGTTLNLGNAPE-----KLALRNEERAIDELKQAIEDKEATTVAIE 55  
Db 219 SDLIEGRGIPPCRNSTKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQAAEATK 278  
Qy 56 AASSDALLALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKQSDNAAS 115  
Db 279 VAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAE 337  
Qy 116 DAWKA--ATPIA-LDVKTKTRPVVKKERQ 145  
Db 338 AEKQAAEATKVAEAEKQAAEATKVAEAEKQ 370

RESULT 13  
US-08-446-137B-2

; Sequence 2, Application US/08446137B  
; Patent No. 6162903  
; GENERAL INFORMATION:  
; APPLICANT: Trowern, Angus R.  
; APPLICANT: Atkinson, Anthony  
; APPLICANT: Murphy, Jonathan P.  
; APPLICANT: Laurence, Oliver S.  
; APPLICANT: Duggieby, Clive J.  
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED  
; FROM L PROTEIN AND THEIR USES  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446.137B  
; FILING DATE: 22-MAY-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McMasters, David D.  
; REGISTRATION NUMBER: 33,963  
; REFERENCE/DOCKET NUMBER: 100084.406  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1027 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-446-137B-2

Query Match 11.9%; Score 92.5; DB 4; Length 1027;  
Best Local Similarity 26.8%; Pred. No. 0.86;

Matches 42; Conservative 23; Mismatches 59; Indels 33; Gaps 6;  
Qy 26 EKALRNE--ERAIDELKQAIEDKE-----ATTAIEA-----ASSDALEALADQTD 71  
Db 226 EKLAARETAKKHIDELKHLADTKELAKRDISATTIDAINDIVARDYMERKLSKET 285  
Qy 72 LQSEAAVVKADNAASDALEALADQTDALQSEAAVVKQSDNAASDAWEKAATPIALDV-- 129  
Db 286 PEPEEVTIKANLIFADG---STQNAEFKGTFAKAVSDAYAYADALKKONGEYTVDVAD 341  
Qy 130 -----KTKDKTPVVKKEERQNVNLTPTTGE 155  
Db 342 KGLTLNKFAGKKEKPEEP--KEEVTIKVNLIFADGK 376

RESULT 14

US-08-968-751-4  
; Sequence 4, Application US/08968751  
; Patent No. 5948643  
; GENERAL INFORMATION:  
; APPLICANT: Rubinfeld, Bonnie  
; APPLICANT: Polakis, Paul G.  
; APPLICANT: Ligenfelter, Carol  
; APPLICANT: Vuong, Terilyn T.  
; TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ONYX Pharmaceuticals, Inc.  
; STREET: 3031 Research Drive  
; CITY: Richmond  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94806  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/968.751  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Giotta, Gregory  
; REGISTRATION NUMBER: 32,028  
; REFERENCE/DOCKET NUMBER: ONYX1024 GG  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 262-8710  
; TELEFAX: (510) 222-9758  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 835 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-968-751-4

Query Match 11.8%; Score 91.5; DB 2; Length 835;  
Best Local Similarity 23.3%; Pred. No. 0.81;  
Matches 37; Conservative 30; Mismatches 73; Indels 19; Gaps 4;

Qy 16 TLTNLLGNA---PEKLALRNEERAIDELKQAIEDKEATTVAIEAASSDALEALADQTDAL 72  
Db 296 TWTALLSRSLQTEKLTVKSQ-ALQE-RDVAIEKQESRVLEQVSAQLECKGQTEQL 353  
Qy 73 QSEAAVVKADNAASDALEALADQTDALQSE-----AEVQSDNNAASDAW 118  
Db 354 ELENRLATDLRAQLQIILANMDSQLKQTHCAQDLAMKDELFCOLTOSNEEQAAQW 413  
Qy 119 EKAATPIALDVKKTKDKTPVVKKEERQNVNLTPTTGEES 157  
Db 414 QREEMALKHMQAELOOQOQAVLAKEVRLDKETLEFADQEN 452



```
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005.069
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/630,822
; FILING DATE: 11-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNELL, GARY J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-17-C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 900 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-005-069-62
```

Query Match 13.0%; Score 101; DB 2; Length 900;

Best Local Similarity 25.3%; Pred. No. 0.11; Mismatches 34; Indels 22; Gaps 5;

```
Qy 17 LTNLLGNAPEKIALRNEERAIDELKKQAIEDKEATTATEAASS---DALEALADOTDALQ 73
Db 492 LEKQINDQEK--LAGELEHNGKLLKQAVELRVASQATEQLNNELOETMGOGLQORDALQ 549
Qy 74 SEEAIVV-----KADNAASD-----ALEALADOTDALQSEAEVYVQSDNAASD--- 116
Db 550 QEVASLQGLKLSRQSSRSOASDMQIELEAKLQALHLELHVRCNECKVTQDNKRLERIS 609
Qy 117 AWEKAATPIALDVKTKDKPKVPVKKERQNVNTLPTTGEESN 158
Db 610 TLEKECASLELELKATQNKYQEVKAHRETEKSRSLVSKEEAN 651
```

## RESULT 8

```
; Sequence 11, Application US/09171517B
; Patent No. 6355254
; GENERAL INFORMATION:
; APPLICANT: Finlay, B. Brett
; APPLICANT: Kenny, Brendan
; APPLICANT: Stein, Markus
; TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN
; FILE REFERENCE: 07422/019001
; CURRENT APPLICATION NUMBER: US/09/171.517B
; CURRENT FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: PCT/CA97/00265
; PRIOR FILING DATE: 1997-04-23
; PRIOR APPLICATION NUMBER: 60/015,999
; PRIOR FILING DATE: 1996-04-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-171-517B-11
```

Query Match 12.3%; Score 95.5; DB 4; Length 312;

Best Local Similarity 27.6%; Pred. No. 0.09; Mismatches 20; Indels 23; Gaps 6;

\*,\*,\*

```
Qy 14 GGTLTNLLGNAPEKIALRNEERAIDEL--KKQAIEDKEATTATEAASS---DALEALAD- 67
Db 102 GGLISSALGILGSPAAAMNAAKGAGEIAEKASSASSKAAAGAAASEVANKALVKATESVADV 161
Qy 68 ---QTDALQSEEAIVVKKADNNAASDALEALADOTDALQSEAEVYVQSDNAASDAWEKAA-- 122
Db 162 ABEASSAMOKAMATTTKAASRASGVADVDYAKATDFAED-----LADAAEKTSTRI 210
Qy 123 TPIALDVKK-TKDTKPVVKKERQNVNTLPTTGEES 157
Db 211 NKLLNSVDKLTNTTAFVAVTSLAEGTKTLPTTISES 246
```

## RESULT 9

```
; US-08-158-353-2
; Sequence 2, Application US/08158353
; Patent No. 5620862
; GENERAL INFORMATION:
; APPLICANT: Padula, Steven J.
; TITLE OF INVENTION: Methods for Diagnosing Early Lyme
; TITLE OF INVENTION: Disease
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,353
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cartoll, Alice O.
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: UCT93-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 212 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-158-353-2
```

Query Match 12.0%; Score 93; DB 1; Length 212;

Best Local Similarity 28.0%; Pred. No. 0.094; Mismatches 46; Conservative 30; Indels 30; Gaps 8;

```
Qy 14 GGTLTNLLGNAPEK---LALRNEE---RAIDELKKQAIEDKEATTATEAASSD--ALEA 64
Db 39 GPNLTEISKKITESNAVVLAVKVEVETLLASIDEVAKKATGNLQNGNLNAGANQNGSLIA 98
Qy 65 -----LADOTDALQSEEAIVVKKADNNAASDALEALADOTDALQSEAEVYVQSDNAASD 116
Db 99 GAYVISTLIAEKLKDLGKNSSELEKIE---DAKCNKRAFTDKLSSHAELGIANGATD 154
Qy 117 AWEKAATPIALDVKTKDKTKPKVPVKKER--QNVNTLPTTGEES 157
Db 155 ANAKAA-----ILKNGTKDKGAQGLEKLFESVKNLSKAAQET 192
```

## RESULT 10



```
Query Match 14.7%; Score 114.5; DB 3; Length 413;
Best Local Similarity 26.2%; Pred. No. 0.0019;
Matches 49; Conservative 15; Mismatches 56; Indels 67; Gaps 6;

Qy 36 AIDELKKQAIED-----KEATTAIE---AASDALEALADQTDALQSEEAHVYK 81
Db 206 AIEILKYGIGYIKLINNKTAEGVTALKDEILASKPAVIDAPELTPAUTTYKLVI-- 263
Qy 82 ADNRASDALEALADQTDALQSEEA-EVVSQDAAADAME----- 119
Db 264 --NGTKLGETTTRAVDAETAKEAFKQYANENGVDGVWYTDATKTFVTETVTEPGDA 321
Qy 120 -----KAATPIALDVKKTKDT-----KPVVKKKEERQNVNTLPT 152
Db 322 PTEPKPEASIPVLPTPATPIAKDDAKKDDTKKDDTKKEDAKKPEAKKEAKKAATLPT 381
Qy 153 TGEESNP 159
Db 382 TGEESNP 388

RESULT 2
US-09-171-517B-6
; Sequence 6, Application US/09171517B
; Patent No. 6355254
; GENERAL INFORMATION:
; APPLICANT: Finlay, B. Brett
; APPLICANT: Kenny, Brendan
; APPLICANT: Stein, Markus
; TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN
; FILE REFERENCE: 07422/019001
; CURRENT APPLICATION NUMBER: US/09/171.517B
; CURRENT FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: PCT/CA97/00265
; PRIOR FILING DATE: 1997-04-23
; PRIOR APPLICATION NUMBER: 60/015,999
; PRIOR FILING DATE: 1996-04-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-171-517B-6

Query Match 13.6%; Score 105.5; DB 4; Length 314;
Best Local Similarity 28.7%; Pred. No. 0.0097;
Matches 45; Conservative 23; Mismatches 68; Indels 21; Gaps 6;

Qy 14 GGTLTNLGNAPKALRNEERAIDELKKQA--TEDKEATT---AIEAASDALEALADQ 68
Db 100 GGAISSVLG-----ILGSFAAINSATKGASDIAQKTASTSSKAIDAASDTATKTLTKA 152
Qy 69 T-----DALOSEEAHVVKADNAASDALEALADQTDALQSEEAHVVSQDAAADAMEKAATP 124
Db 153 TESVADAVEDASSVMOQAMTTATRAASRTSDVADDIADSAQRASOLAENAAADAQKASRA 212
Qy 125 ----IALDVKKTKDTPVVKKEERQNVNTLPTTGEES 157
Db 213 SRFMAAVD-KITGSTPFFIAVTSLAEGTKTLPTTVSES 248

RESULT 3
US-09-171-517B-9
; Sequence 9, Application US/09171517B
; Patent No. 6355254
; GENERAL INFORMATION:
; APPLICANT: Finlay, B. Brett
; APPLICANT: Kenny, Brendan
; APPLICANT: Stein, Markus
; TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN
; FILE REFERENCE: 07422/019001
```

```
; CURRENT APPLICATION NUMBER: US/09/171.517B
; CURRENT FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: PCT/CA97/00265
; PRIOR FILING DATE: 1997-04-23
; PRIOR APPLICATION NUMBER: 60/015,999
; PRIOR FILING DATE: 1996-04-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-171-517B-9

Query Match 13.6%; Score 105.5; DB 4; Length 314;
Best Local Similarity 28.7%; Pred. No. 0.0097;
Matches 45; Conservative 23; Mismatches 68; Indels 21; Gaps 6;

Qy 14 GGTLTNLGNAPKALRNEERAIDELKKQA--TEDKEATT---AIEAASDALEALADQ 68
Db 100 GGAISSVLG-----ILGSFAAINSATKGASDIAQKTASTSSKAIDAASDTATKTLTKA 152
Qy 69 T-----DALOSEEAHVVKADNAASDALEALADQTDALQSEEAHVVSQDAAADAMEKAATP 124
Db 153 TESVADAVEDASSVMOQAMTTATRAASRTSDVADDIADSAQRASOLAENAAADAQKASRA 212
Qy 125 ----IALDVKKTKDTPVVKKEERQNVNTLPTTGEES 157
Db 213 SRFMAAVD-KITGSTPFFIAVTSLAEGTKTLPTTVSES 248

RESULT 4
US-09-171-517B-10
; Sequence 10, Application US/09171517B
; Patent No. 6355254
; GENERAL INFORMATION:
; APPLICANT: Finlay, B. Brett
; APPLICANT: Kenny, Brendan
; APPLICANT: Stein, Markus
; TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN
; FILE REFERENCE: 07422/019001
; CURRENT APPLICATION NUMBER: US/09/171.517B
; CURRENT FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: PCT/CA97/00265
; PRIOR FILING DATE: 1997-04-23
; PRIOR APPLICATION NUMBER: 60/015,999
; PRIOR FILING DATE: 1996-04-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-171-517B-10

Query Match 13.6%; Score 105.5; DB 4; Length 314;
Best Local Similarity 28.7%; Pred. No. 0.0097;
Matches 45; Conservative 23; Mismatches 68; Indels 21; Gaps 6;

Qy 14 GGTLTNLGNAPKALRNEERAIDELKKQA--TEDKEATT---AIEAASDALEALADQ 68
Db 100 GGAISSVLG-----ILGSFAAINSATKGASDIAQKTASTSSKAIDAASDTATKTLTKA 152
Qy 69 T-----DALOSEEAHVVKADNAASDALEALADQTDALQSEEAHVVSQDAAADAMEKAATP 124
Db 153 TESVADAVEDASSVMOQAMTTATRAASRTSDVADDIADSAQRASOLAENAAADAQKASRA 212
Qy 125 ----IALDVKKTKDTPVVKKEERQNVNTLPTTGEES 157
Db 213 SRFMAAVD-KITGSTPFFIAVTSLAEGTKTLPTTVSES 248

RESULT 5
US-09-171-517B-11
; Sequence 11, Application US/09171517B
; Patent No. 6355254
; GENERAL INFORMATION:
; APPLICANT: Finlay, B. Brett
; APPLICANT: Kenny, Brendan
; APPLICANT: Stein, Markus
; TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN
; FILE REFERENCE: 07422/019001
```

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: October 13, 2002, 04:45:42 ; Search time 24.6578 Seconds  
(without alignments)  
157.503 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQRIIPNGOTLTNL.....KKEERQNVNLTPTTGESNP 159

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114.5	14.7	413	3	US-08-669-408B-10
2	105.5	13.6	314	4	US-09-171-517B-6
3	105.5	13.6	314	4	US-09-171-517B-9
4	105.5	13.6	314	4	US-09-171-517B-10
5	105	13.5	654	3	US-08-669-408B-2
6	101	13.0	900	2	US-08-630-822A-62
7	101	13.0	900	2	US-09-005-069-62
8	95.5	12.3	312	4	US-09-171-517B-11
9	93	12.0	212	1	US-08-158-353-2
10	93	12.0	443	2	US-08-795-475-6
11	92.5	11.9	472	2	US-08-216-894-10
12	92.5	11.9	472	4	US-09-115-746-10
13	92.5	11.9	1027	4	US-08-446-137B-2
14	91.5	11.8	835	2	US-08-968-751-4
15	91	11.7	593	2	US-08-591-079-8
16	91	11.7	593	2	US-08-591-079-10
17	90.5	11.6	516	2	US-08-762-106-8
18	90.5	11.6	516	3	US-08-745-404-2
19	90.5	11.6	516	4	US-09-320-774-8
20	90.5	11.6	527	2	US-08-762-106-9
21	90.5	11.6	527	4	US-09-320-774-9
22	90.5	11.6	552	3	US-08-745-404-3
23	90	11.6	190	1	US-08-106-981-2
24	90	11.6	619	1	US-08-465-746-2
25	90	11.6	619	1	US-08-214-164-2
26	90	11.6	619	2	US-08-467-852A-3
27	90	11.6	619	2	US-08-246-636-2

28	90	11.6	619	2	US-08-247-491A-3	Sequence 3, Appl1
29	90	11.6	619	2	US-08-319-795-2	Sequence 2, Appl1
30	90	11.6	619	2	US-08-468-985-2	Sequence 2, Appl1
31	90	11.6	619	3	US-08-312-949-2	Sequence 2, Appl1
32	90	11.6	648	1	US-08-072-070-2	Sequence 2, Appl1
33	90	11.6	648	1	US-08-469-434-2	Sequence 2, Appl1
34	90	11.6	648	1	US-08-214-222-2	Sequence 2, Appl1
35	90	11.6	648	2	US-08-467-852A-2	Sequence 2, Appl1
36	90	11.6	648	2	US-08-468-718-2	Sequence 2, Appl1
37	90	11.6	648	2	US-08-247-491A-2	Sequence 2, Appl1
38	90	11.6	648	3	US-08-446-201-3	Sequence 3, Appl1
39	90	11.6	695	1	US-08-127-499A-23	Sequence 23, Appl1
40	90	11.6	695	1	US-08-482-847-23	Sequence 23, Appl1
41	89	11.5	289	1	US-08-072-070-4	Sequence 4, Appl1
42	89	11.5	289	1	US-08-469-434-4	Sequence 4, Appl1
43	89	11.5	289	1	US-08-214-222-4	Sequence 4, Appl1
44	89	11.5	289	2	US-08-467-852A-5	Sequence 5, Appl1
45	89	11.5	289	2	US-08-468-718-4	Sequence 4, Appl1

## ALIGNMENTS

RESULT 1  
US-08-669-408B-10  
; Sequence 10, Application US/08669408B  
; Patent No. 6100055  
; GENERAL INFORMATION:  
; APPLICANT: GUSS, Bengt  
; APPLICANT: JONSSON, Hans  
; APPLICANT: LINDBERG, Martin  
; APPLICANT: MUELLER, Hans-Peter  
; APPLICANT: RANTAMAKI, Liisa K.  
; TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING  
; TITLE OF INVENTION: PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/669,408B  
; FILING DATE: 03-JUL-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/SE94/00826  
; FILING DATE: 06-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: SE 9302855-3  
; FILING DATE: 06-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 61743/102  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 413 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-669-408B-10



**THIS PAGE BLANK (uspto)**

Db 311 AAGGCTACTGAGGCTGCTGAGGATTCTGAGGCTGATGAGGCTGCTGAGGTTTTGAAGGAT 252  
Qy 129 ValLysLysThrLysAspThrLys 136  
Db 251 ACTGAGGCTGCTGAGGATTCTGAG 228

Search completed: October 13, 2002, 05:37:48  
Job time : 1854.09 secs



```
QY 29 AlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLys 48
Db 969 GATGCAGCTAACCTGAAAAA-----GAAAGCTCCAAATCACAGCA----- 1010
QY 49 GluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGln 68
Db 1011 -----GCAGCCCTAGAAAAACAACACTAGAGCAACTAAAAAAGAG 1049
QY 69 ThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAsp 88
Db 1050 TTAGCTGATTACAGGCTTAATAGCAGCAACCAACCAAGAAAAAGAAAGTTAGAAAGCT 1109
QY 89 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValVal 108
Db 1110 GAAGCAAAAGCTCTTAAAGAGCAATTA---ACAAACAAGCTGAAGAGCTGTGAAGCTA 1166
QY 109 GlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAsp 128
Db 1167 AAAGCAGATAAAGCTTCAGCA-----GCTCAAAAACCA-----GAT 1202
QY 129 ValLysLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsnValAsn 148
Db 1203 ACCAAAGCTGGCAATAAAGAGGTTCCCAACAGACCGTCACAAACAAGACCAACACTAAT 1262
QY 149 Thr-----LeuProThrThrGlyGluGlu---SerAsn 158
Db 1263 ACAGCTCTATGGCGCAACAAGACAGACATTAACCGTCAACAGCGCAAGAAACCAAC 1322
QY 159 Pro 159
Db 1323 CCA 1325

RESULT 38
SG148IGG 1469 bp DNA linear BCT 12-JUL-1995
LOCUS Streptococcus G 148 gene fragment for IgG-binding protein G.
DEFINITION X04015
ACCESSION X04015
VERSION X04015.1 GI:47071
KEYWORDS gamma-immunoglobulin binding protein G.
SOURCE Streptococcus sp.
ORGANISM Streptococcus sp.
REFERENCE 1 (bases 1 to 1469)
AUTHORS Guss B., Eliasson M., Olsson M., Uhlen, M., Frej A.K., Jornvall, H.,
Flock, J.I. and Lindberg, M.
TITLE Structure of the IgG-binding regions of streptococcal protein G
JOURNAL EMBO J. 5 (7), 1567-1575 (1986)
MEDLINE 86300657
COMMENT Repetitive structures of 24 amino acids, A1, A2 and A3, are found
interrupted by two unrelated repeats of 51 amino acids, B1 and B2.
Further downstream, following a 12-residue unique 'spacer'-region'
(5), there is a stretch of 55 amino acids repeated three times, C1,
C2 and C3. The latter regions are interrupted by two 15-residue
regions, called D1 and D2. Directly following these repeats there
is a region W, which consists of a repetitive, extremely
hydrophilic, structure containing several proline residues. This
region resembles a similar structure in the streptococcal M protein
and in protein A, which is thought to mediate the binding of the
protein to the cell wall.
Location/Qualifiers
1..1469
/organism="Streptococcus sp."
/strain="G 148"
/db_xref="taxon:1306"
<1..1443
/codon_start=1
/product="protein G"
/protein_id="CAA27638.1"
/db_xref="GI:47072"
/db_xref="SWISS-PROT:P19909"
/translation="EFNKYGVSDYKLNINNAKTVEGVKDLQAVVESAKARISEAT"

FEATURES
source
CDS
```

```
BASE COUNT 546 a 279 c 297 g 347 t
ORIGIN
Alignment Scores:
Pred. No.: 2.68 Length: 1469
Score: 112.00 Matches: 44
Percent Similarity: 39.44% Conservative: 12
Best Local Similarity: 30.99% Mismatches: 62
Query Match: 14.41% Indels: 24
DB: 1 Caps: 5

US-09-847-539A-6 (1-159) x SG148IGG (1-1469)
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGlu-----Ala 56
Db 943 AAACCAGAGTGATCGTCTGCTGAATTAACACCGCGTGACAACTTACAAACTTGT 1002
QY 57 AlaSerSerAspAlaLeu-----GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 74
Db 1003 ATTAATCGTAAACATTGAAGCGCAACAACTACTAAAGCAGTAGACGAGAACTGCA 1062
QY 75 GluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAla 94
Db 1063 GAAAAAGCCCTCAACAACAACTACGCTAACGCAACCGGTGTGATGCTTTGGACTATGAT 1122
QY 95 AspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAla 114
Db 1123 GATCGGACTAAGACCTTTACGGTAACCTGAAATG---GTTACAGAGGTTCCCTGGTGATCA 1179
QY 115 SerAspAlaTrpGluLys-----AlaAlaThr 123
Db 1180 CCAACTGAACCAAGAAACCAAGCAAGTATCCCTCTTGTTCCTGTTAACTCTCGCAACT 1239
QY 124 ProIleAlaLeuAspValLysLysThrLysAspThr-----LysPro 137
Db 1240 CCAATTCCTAAAGATGACGCTAACAGAACAGCATACTAAGAAGAAGATGCTTAAAAACCA 1299
QY 138 ValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSer 157
Db 1300 GAAGCTAAGAAAGAAGACGCTAAGAAAGCTGAAAGCTCTCTCTACAACTGCTGAAGGAAGC 1359
QY 158 AsnPro 159
Db 1360 AACCCA 1365

RESULT 39
LOCUS A12446
DEFINITION 1.5 kb EcoRI/HindIII restriction fragment of pSPG1.
ACCESSION A12446
VERSION A12446.1 GI:491352
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 1469)
AUTHORS
TITLE
JOURNAL
FEATURES
source
METHOD AND MEANS FOR PRODUCING A PROTEIN HAVING THE SAME IGC
SPECIFICITY AS PROTEIN G
Patent: WO 8705631-A 4 24-SEP-1987;
Location/Qualifiers
1..1469
/organism="synthetic construct"
/db_xref="taxon:32630"
1..1443
CDS
```

```
Alignment Scores:
Pred. No.: 256 Length: 146491
Score: 113.00 Matches: 37
Percent Similarity: 47.88% Conservativity: 42
Best Local Similarity: 22.42% Mismatches: 60
Query Match: 14.54% Indels: 26
DB: 2 Gaps: 6

US-09-847-539A-6 (1-159) x AC019950 (1-146491)
Qy 11 IleProAsnGlyClyThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeuAlaLeu 30
    :::::::::::
Db 104506 GTCCCCAAT-----ATAATAGTCAGCCCGCAGAAATGGATGG 104468
    :::::::::::
Qy 31 ArgAsnGlu-----GluArgAlaIleAspGluLeuLysLysGlnAla 44
    :::::::::::
Db 104467 GACTCAGAAGCGCGGACGAGAGGCTTCGAAGCGCGAGTCCAGATCAAGAAGAGGCT 104408
    :::::::::::
Qy 45 IleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSer-----SerAsp 60
    :::::::::::
Db 104407 GAGCTTGAAACGATCAAGAGGAATTTCAAGAGGATTCACCAATCCGATCCGAT 104348
    :::::::::::
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
    :::::::::::
Db 104347 TCGGTGGCGGAATTAAAGACAGCGGTAACCTGCTCTGAGGCGGAGGATGATGTCATCTA 104288
    :::::::::::
Qy 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
    :::::::::::
Db 104287 GAGGCGGATAACATACGCAAGAGCTGCTAGATGATTAATGTCGCGAGCGGAG---CTA 104231
    :::::::::::
Qy 101 GlnSerGluGluAlaGluValValGlnSer---AspAsnAlaAlaSerAspAlaThr 119
    :::::::::::
Db 104230 CTGTATCAGAGGAAGACATTTGTTCAATCTGAAGAGATGCCAGACAGCGGCTAGAT 104171
    :::::::::::
Qy 120 LysAlaAlaThr-----ProIleAlaLeuAspValLysLysThr 132
    :::::::::::
Db 104170 AGGTCAGTACCGACGACGAGTACCTGGTGGCCACCTACCGAGTATCTACCGAACAATG 104111
    :::::::::::
Qy 133 LysAspThrLysProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThr 152
    :::::::::::
Db 104110 GAGATTGATGAACGACGAGCGGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 104051
    :::::::::::
Qy 153 ThrGlyGluGluSer 157
    :::::::::::
Db 104050 GAGCAGAGGAGGAGGCC 104036

RESULT 37
SPU52008
LOCUS
DEFINITION
Streptococcus pyogenes emm gene cluster, mrp50, emm50 and enn50
genes, complete cds.
ACCESSION
U52008
VERSION
U52008.1 GI:1354360
KEYWORDS
Streptococcus pyogenes strain-B514.
SOURCE
Streptococcus pyogenes
ORGANISM
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 4469)
AUTHORS
Yung,D.L. and Hollingshead,S.K.
TITLE
DNA sequencing and gene expression of the emm gene cluster in an
M50 group A streptococcus strain virulent for mice
JOURNAL
Infect. Immun. 64 (6), 2193-2200 (1996)
MEDLINE
96239037
REFERENCE
2 (bases 1 to 4469)
AUTHORS
Hollingshead,S.K. and Yung,D.L.
TITLE
Direct Submission
JOURNAL
Submitted (21-MAR-1996) Susan K. Hollingshead, Microbiology, Univ.
of Alabama at Birmingham, 845 19th Street South BBRB 406/6,
Birmingham, AL 35294, USA
FEATURES
Location/Qualifiers
1..4469
/organism="Streptococcus pyogenes"
```

```
/strain="B514"
/db_xref="taxon:1314"
123..1400
/gene="mrp50"
123..1400
/gene="mrp50"
/note="IgG binding protein; Allele: 50; M-related protein"
/codon_start=1
/transl_table=11
/product="Mrp50"
/db_xref="GI:1354361"
/translation="MSKRNPKHYSLRKLTGTASVAVALTVLGTGLNTTDDVKAETV
GRFDEQVRKAREKAIEDVFGYTGARSYVGSNLPNRLTPTKSLKLMQMYKTELQK
KEEDTLTKALTDITTEKRIESENAYKKGELGOLKAAAEAEAKDALNNKQKISDLT
NENALQKEATIEGYVOTIQNASREIAAKQELAAAKSQAELNAETALKQOQDSKTEE
TAKQSEATLENLGSAKHELTDLQAKLDATATKAKLESQVTTLENLGSAKREL
DLOAKLDARNAEKELQSOAALEKQLEATKELADLQAKLAATNOEKLEAEAKAL
KEQLTQAEELAKLADKASGAKPDTPGNGKEVTPRSOTRINTNTAPMAQTKRQLP
STGFTTTPFFTAALTVIASAGVLAKRKEEN"
1625..2866
/gene="emm50"
1625..2866
/gene="emm50"
/note="IgG binding protein; Allele: 50; M protein of
serotype 50"
/codon_start=1
/transl_table=11
/product="Emm50"
/protein_id="AAB05613.1"
/db_xref="GI:1354362"
/translation="MARKDTNLQYSLRKLKTGTASVAVAVAVLGVAGFANQTTVKAEEA
GASRTITSENTISKLYDENSLIEERADLLGLEKEDKLESVERHYLRTLDQYKEH
QKEQOQOQOEEROKYQOLEROQREVEKRYQOLOQOQOLEKEKOISASRKSLS
RDLEASRAKKELEAEHQKLEKEKQISASROGLSDLEASREAKKYEADLAALTA
HQKLEEKQISASRQGLSDLEASREAKKRYERDLAALTAHQKLEKEKQISDSARQ
GLSDLEASREAKKRYERDLAALTAHQKLEKEKQISDSARQ
AKALKQELAKQAEELAKLGNQTPNAKVAPOANRSRSAMTQOKRTLPSTGETANPFFT
BEATVTVVSAGMLAKRKEEN"
3077..4204
/gene="enn50"
3077..4204
/gene="enn50"
/note="IgA binding protein; Allele: 50; M-related protein"
/codon_start=1
/transl_table=11
/product="Enn50"
/protein_id="AAB05614.1"
/db_xref="GI:1354363"
/translation="MARQTKTKYSLRKLKTGTASVAVALTVLGVAGFANQTEVRADQA
ASGTSVNGASTSVESFYDALRGENADLRNVNAYLEKINAEKKNKLEATNKELN
ENYKLOPDIDALEKEKEDSKTLAKTTKENEI SPASRGLSRDLSEASRAAKKELEAK
HOKLEAKNKLTEANQISASRKGISNDLEASRAKKELEAKHOKLEADHQALEAKHQ
KLAGEHVSFTRKGLSDLEASREANKKVTSELTOAKQISALSESKEKLSKEKAEEL
QAKLDAQKALKQELAKQAEELAKLGNQTPNAKVAPOANRSRSAMTQOKRTLPSTGETANP
NKGRSOLPSTGEANPFTTAAATVTVVSAGMLTKRKEEN"
BASE COUNT 1841 a 860 c 870 g 898 t
ORIGIN
Alignment Scores:
Pred. No.: 7.72 Length: 4469
Score: 112.50 Matches: 49
Percent Similarity: 42.24% Conservativity: 19
Best Local Similarity: 30.43% Mismatches: 54
Query Match: 14.48% Indels: 39
DB: 1 Gaps: 8

US-09-847-539A-6 (1-159) x SPU52008 (1-4469)
Qy 16 ThrLeuThrAsnLeuLeuGlyAsnAlaProGlu-----LysLeu 28
    :::::::::::
Db 909 ACCTTAGAAAACCTCCTAGGTTCAAGCTAGAGTGAATTCAGTCTGCAAGCTAAGCTA 968
    :::::::::::
```









```

* 199307 201611: contig of 2305 bp in length
* 201612 201631: gap of unknown length
* 201632 202744: contig of 1113 bp in length
* 202745 202764: gap of unknown length
* 202765 204092: contig of 1328 bp in length
* 204093 204112: gap of unknown length
* 204113 205425: contig of 1313 bp in length
* 205426 205445: gap of unknown length
* 205446 206650: contig of 1205 bp in length
* 206651 206670: gap of unknown length
* 206671 207282: contig of 612 bp in length
* 207283 207302: gap of unknown length
* 207303 208250: contig of 948 bp in length
* 208251 208270: gap of unknown length
* 208271 209082: contig of 812 bp in length
* 209083 209102: gap of unknown length
* 209103 210487: contig of 1385 bp in length
* 210488 210507: gap of unknown length
* 210508 210791: contig of 284 bp in length.
FEATURES
    source
        1..210791
            Location/Qualifiers
                organism="Mus musculus"
                strain="C57BL6/J"
                db_xref="taxon:10090"
                chromosome="15"
                clone="RP23-384L8"
                sex="male"
            1..21856
                /note="assembly_name:Contig134"
            21877..48688
                /note="assembly_name:Contig133"
            48709..65965
                /note="assembly_name:Contig132"
            65986..80051
                /note="assembly_name:Contig131"
            80072..94631
                /note="assembly_name:Contig130"
            94652..108022
                /note="assembly_name:Contig129"
            108043..119442
                /note="assembly_name:Contig128"
            119463..132361
                /note="assembly_name:Contig127"
            132382..141581
                /note="assembly_name:Contig126"
            clone_end:Sp6
            vector_side:left"
            141602..151955
                /note="assembly_name:Contig125"
            151976..159593
                /note="assembly_name:Contig124"
            159614..167366
                /note="assembly_name:Contig123"
            167387..173218
                /note="assembly_name:Contig122"
            173239..179907
                /note="assembly_name:Contig121"
            179928..185771
                /note="assembly_name:Contig120"
            185792..189317
                /note="assembly_name:Contig119"
            189338..193220
                /note="assembly_name:Contig118"
            193241..196107
                /note="assembly_name:Contig117"
            196128..199286
                /note="assembly_name:Contig116"
            199307..201611
                /note="assembly_name:Contig115"
            201632..202744
                /note="assembly_name:Contig114"
            clone_end:T7
            vector_side:left"

```

```

misc_feature 202765..204092
              /note="assembly_name:Contig113"
misc_feature 204113..205425
              /note="assembly_name:Contig112"
misc_feature 205446..206650
              /note="assembly_name:Contig111"
misc_feature 206671..207282
              /note="assembly_name:Contig110"
misc_feature 207303..208250
              /note="assembly_name:Contig109"
misc_feature 208271..209082
              /note="assembly_name:Contig108"
misc_feature 209103..210487
              /note="assembly_name:Contig107"
misc_feature 210508..210791
              /note="assembly_name:Contig106"
BASE COUNT 52816 a 53362 c 53372 g 50609 t 632 others
ORIGIN

```

```

Alignment Scores:
Pred. No.:      285      Length:      210791
Score:          114.50    Matches:      36
Percent Similarity: 50.81%  Conservative: 27
Best Local Similarity: 29.03%  Mismatches: 50
Query Match:      14.74%    Indels:      11
DB:               2        Gaps:       3
US-09-847-539A-6 (1-159) x AC023354 (1-210791)

```

```

QY 26 GluLysLeuAlaLeuArgAsnGluGluArgAlaLeuAlaLeuLysGlnAlaLeu 45
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109068 GAGCACCTACGGGTCGGCCAGAGAGAGGATGCAGCAGGAGGAGGAGGAGGAGTA 109127
QY 46 GluAspLysGluAla-----TThrAlaLeuGluAlaLeuAlaSer 58
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109128 CAAGAAGTCAGATCTGTCATGCTATATACCTCATAGCGCAGAGGAGGAGGAGG 109187
QY 59 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109188 GCAGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 109238
QY 79 ValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109239 GCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 109298
QY 99 AlaLeuGlnSerGluGluAlaGluValGlnSerAspAsnAlaAlaSerAspAlaTrp 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109299 GCA---GAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 109355
QY 119 GluLysAlaAlaThrProileAlaLeuAspValLysLysThrLysAspThrLysProVal 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109356 GGATCACTACAAATTCACGTCAGTCTTATGTACAAACATAATCACTACATAAGAGTGTC 109415
QY 139 ValLysLysGlu 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109416 TCAAGGAAGCAG 109427
RESULT 34
AC104834/c 222304 bp DNA linear HTG 25-JAN-2002
LOCUS Mus musculus chromosome 15 clone RP23-274F23 strain C57BL6/J,
DEFINITION WORKING DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION AC104834
VERSION AC104834.3 GI:18030119
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 222304)
REFERENCE Montgomery,K.T., Grills,G., Li,L., Brown,W.A., Chiu,D., Decker,J.,
Fusina,M., Halder,A., Kellier,A., Perera,A., Shm,C., Thomas,E.,
Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.

```

EADGADAIKRAQAEADAIKLRGEALRQNPVMELEAIKWNKGOLPOYMTGANTPPE  
TALK"  
complement(2427..2459)  
/gene="HCM2.0006c"  
/note="PS00013 Prokaryotic membrane lipoprotein lipid  
attachment site"  
complement(2493..2693)  
/gene="HCM2.0007c"  
complement(2493..2693)  
/gene="HCM2.0007c"  
/note="HCM2.0007c", possible membrane protein, len: 66 aa;  
highly similar to TR:CAB62372 (EMBL:AL117211), YPM1.49AC,  
versinia pestis CO-92 hypothetical protein from plasmid  
pWT1 (66 aa), fasta scores; E(): 3.5e-25, 97.0% identity  
in 66 aa overlap. Contains hydrophobic, possible

## Alignment Scores:

Pred. No.: 141 Length: 106516  
Score: 114.50 Matches: 40  
Percent Similarity: 46.51% Conservatives: 20  
Best Local Similarity: 31.01% Mismatches: 44  
Query Match: 14.74% Indels: 25  
DB: 1 Gaps: 5

US-09-847-539a-6 (1-159) x STVPPHCM2 (1-106516)

QY 37 ILASpGluLeuLysGlnAlaLeuGluAspLysGluAlaThrThrAlaLeuGluAla 56  
:||||| :||||| :||||| :||| :|||||  
Db 39784 GTCGATACCATTCAGTCGGAGTCTCTCAAGGCTCTGGCTSCAGCCGCCAGATCGGAAGCA 39725  
:||||| :||||| :||||| :||||| :|||||  
QY 57 ALASerSerAspAlaLeu-----GluAlaLeu 65  
:||||| :||||| :||||| :||||| :|||||  
Db 39724 CGCGTGCAGAGCGGTAACTCTAAACAGACGCTCGCGGCGAGCGAAGCTAATGCTAAG 39665  
:||||| :||||| :||||| :||||| :|||||  
QY 66 ALASpGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAlaAspAsnAla 85  
:||||| :||||| :||||| :||||| :|||||  
Db 39664 GCAAGTGAAACAGCTGCAGCTGCCTCTCAGCAGCGAGCGCTACCAGTGAAGTAAGGCC 39605  
:||||| :||||| :||||| :||||| :|||||  
QY 86 ALASerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAla 105  
:||||| :||||| :||||| :||||| :|||||  
Db 39604 CGGCAAGCAAGAGAGCTCGCGGAGCATCACAAACAGCTCATTAACAGTACAGAGGTT 39545  
:||||| :||||| :||||| :||||| :|||||  
QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaThrGluLysAlaAlaThrProIle 125  
:||||| :||||| :||||| :||||| :|||||  
Db 39544 GCAGCGCCCATCGACACTCGCGGAGAGTGAAGCA---GAAAGGCG----- 39500  
:||||| :||||| :||||| :||||| :|||||  
QY 126 ALASpValLysLys-----ThrLysAspThrLysProValValLysLysGlu 142  
:||||| :||||| :||||| :||||| :|||||  
Db 39499 -----AAGAAATTTGCCGATCTCTTAGATGTAACAACAGCTGCTGCATATAAGAT 39452  
:||||| :||||| :||||| :||||| :|||||  
QY 143 GluArgGlnAsnValAsnThrLeuPro 151  
:||||| :||||| :||||| :||||| :|||||  
Db 39451 -----CAGAACCTTGCTGCACCTGCCA 39431  
:||||| :||||| :||||| :||||| :|||||  
RESULT 33  
AC023354  
LOCUS  
DEFINITION Mus musculus chromosome 15 clone RP23-384L8 strain C57BL6/J,  
WORKING DRAFT SEQUENCE, 29 unordered pieces.  
ACCESSION AC023354  
VERSION AC023354.6 GI:18092970  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 210791)  
REFERENCE  
AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,  
Gordon,M., Goltz,J.S. and Kucherlapati,R.  
TITLE High Throughput Mouse Sequencing  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 210791)

## AUTHORS

Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,  
Gordon,M., Goltz,J.S. and Kucherlapati,R.

TITLE  
JOURNAL

Direct Submission  
Submitted (14-FEB-2000) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA

## COMMENT

On Jan 9, 2002 this sequence version replaced gi:11094415.  
-----Genome Center  
Center: Harvard Partners Genome Center  
Center Code: HPGC  
Web site:  
http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mous  
e.html  
Contact: gkntm@ceped.bwh.harvard.edu  
-----Summary Statistics

Center project name: ABA  
Sequencing vector: pUC18; L08752  
Chemistry: Dye-terminator Big Dye; 100%  
\*Consensus quality: 198418 at least Q20  
\*Consensus quality: 195838 at least Q30  
\*Consensus quality: 191563 at least Q40  
Estimated insert size: agarose-PP - N/A  
\*\*Estimated insert size: 210231 - sum-of-contigs  
Quality coverage: agarose-PP - N/A  
Quality coverage: 4.8 x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 29 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 21856: contig of 21856 bp in length  
\* 21857 21876: gap of unknown length  
\* 21877 48688: contig of 26812 bp in length  
\* 48689 48708: gap of unknown length  
\* 48709 65965: contig of 17257 bp in length  
\* 65966 65985: gap of unknown length  
\* 65986 80051: contig of 14066 bp in length  
\* 80052 80071: gap of unknown length  
\* 80072 94631: contig of 14560 bp in length  
\* 94632 94651: gap of unknown length  
\* 94652 108022: contig of 13371 bp in length  
\* 108023 108042: gap of unknown length  
\* 108043 119443: contig of 11400 bp in length  
\* 119443 119462: gap of unknown length  
\* 119463 132361: contig of 12899 bp in length  
\* 132362 132381: gap of unknown length  
\* 132382 141581: contig of 9200 bp in length  
\* 141582 141601: gap of unknown length  
\* 141602 151955: contig of 10354 bp in length  
\* 151956 151975: gap of unknown length  
\* 151976 159593: contig of 7618 bp in length  
\* 159594 159613: gap of unknown length  
\* 159614 167366: contig of 7753 bp in length  
\* 167367 167386: gap of unknown length  
\* 167387 173218: contig of 5832 bp in length  
\* 173219 173238: gap of unknown length  
\* 173239 179907: contig of 6669 bp in length  
\* 179908 179927: gap of unknown length  
\* 179928 185771: contig of 5844 bp in length  
\* 185772 185791: gap of unknown length  
\* 185792 189317: contig of 3526 bp in length  
\* 189318 189337: gap of unknown length  
\* 189338 193220: contig of 3883 bp in length  
\* 193221 193240: gap of unknown length  
\* 193241 196107: contig of 2867 bp in length  
\* 196108 196127: gap of unknown length  
\* 196128 199286: contig of 3159 bp in length  
\* 199287 199306: gap of unknown length

```

/misc_feature
/db_xref="taxon:90370"
complement(1..29)
/gene="HCM2.0001c"
CDS
complement(1..29)
/partial
/gene="HCM2.0001c"
/note="HCM2.0001c, hypothetical protein, len: 68 aa; highly similar to TR:CAB55227 (EMBL:AL117211), YPWT1.45C, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (71 aa), fasta scores: E(): 5.9e-25, 93.3% identity in 60 aa overlap and TR:O68760 (EMBL:AF074611), Y1119, Yersinia pestis K1M5 hypothetical protein from plasmid pMT1 (71 aa), fasta scores: E(): 5.9e-25, 93.3% identity in 60 aa overlap. Spans the end of the sequence"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09868.1"
/db_xref="GI:16505982"
/translation="MAVTLAGLE"
1..232
/note="95.7% identical to Yersinia pestis CO-92 plasmid pMT1 (EMBL:AL117211) bases 48304..48535"
complement(29..334)
/gene="HCM2.0002c"
CDS
complement(29..334)
/gene="HCM2.0002c"
/note="HCM2.0002c, possible periplasmic protein, len: 101 aa; similar to TR:CAB55228 (EMBL:AL117211), YPWT1.46C, Yersinia pestis hypothetical protein from plasmid pMT1 (111 aa), fasta scores: E(): 2.8e-26, 95.5% identity in 67 aa overlap. The N-terminal 37 aa are highly similar to YPWT1.46AC (EMBL:AL117211) Yersinia pestis hypothetical protein from plasmid pMT1 (59 aa) (86.8% identity in 38 aa overlap) which lies immediately upstream of YPWT1.46C in pMT1. Contains a possible N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAD09869.1"
/db_xref="GI:16505983"
/translation="MKTVLTLLIVSIAFVLCIFGTATRKALAPVNGAMVYDSGQ FYVVRDRLVSSPLTKRNGKLSGVGDREYFNKSGVGVGHAKNASYLPDDCKEVG"
complement(36..43)
/gene="HCM2.0002c"
/note="possible RBS"
224..6722
/note="96.5% identical to Yersinia pestis CO-92 plasmid pMT1 (EMBL:AL117211) bases 48733..55208"
complement(375..650)
/gene="HCM2.0003c"
CDS
complement(375..650)
/gene="HCM2.0003c"
/note="HCM2.0003c, hypothetical protein, len: 91 aa; highly similar to TR:CAB62371 (EMBL:AL117211), YPWT1.46BC, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (91 aa), fasta scores: E(): 0, 94.5% identity in 91 aa overlap and TR:O92GX9 (EMBL:AF074611), Y1117, Yersinia pestis K1M5 hypothetical protein from plasmid pMT1 (91 aa), fasta scores: E(): 0, 94.5% identity in 91 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09870.1"
/db_xref="GI:16505984"
/translation="MKILVRISATDYDVYPLFMVKCDGLNDEFIQAIERNLVEYTG MDADVYVDDGVGVHNGSCWYVDDTMTVSDDEAAHLERILGISTFE"
complement(658..663)
/note="possible RBS"
complement(719..1129)
/gene="HCM2.0004c"
CDS
complement(719..1129)

```

---

```

/gene="HCM2.0004c"
/note="HCM2.0004c, hypothetical protein, len: 136 aa; highly similar to TR:CAB55229 (EMBL:AL117211), YPWT1.47C, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (136 aa), fasta scores: E(): 0, 97.8% identity in 136 aa overlap and TR:O68762 (EMBL:AF074611), Y1116, Yersinia pestis K1M5 hypothetical protein from plasmid pMT1 (136 aa), fasta scores: E(): 0, 97.8% identity in 136 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09871.1"
/db_xref="GI:16505985"
/translation="MKLYDAANAALDVVDTEIAOQFPPEMATQLRALAEANNAPEP SEEDADQRFIRMAEEVGPTPTAEQAMLLKYFKEAGENLPVDDTPHFWHAARWKFQV IYTRGSGKDMVYVHLMHIDKAVDRTLEKFFPPA"
complement(1113..1484)
/gene="HCM2.0005c"
CDS
complement(1113..1484)
/gene="HCM2.0005c"
/note="HCM2.0005c, hypothetical protein, len: 123 aa; similar to TR:Q38665 (EMBL:X78401), ninX, Bacteriophage P22 NinX protein (112 aa), fasta scores: E(): 8.5e-15, 45.6% identity in 114 aa overlap. Highly similar to TR:CAB55230 (EMBL:AL117211), YPWT1.48C, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (123 aa), fasta scores: E(): 0, 87.8% identity in 123 aa overlap and TR:Q9ZGY0 (EMBL:AF074611), Y1115, Yersinia pestis K1M5 hypothetical protein from plasmid pMT1 (123 aa), fasta scores: E(): 0, 87.8% identity in 123 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09872.1"
/db_xref="GI:16505986"
/translation="MKNYSMTDFEINCLVAETGHRPLISQYGWGQSGDYTAVVA IGPNGAGTFCNCDPEDADWDIIYRHRIGVIPARQPGFWAAHRKVDSTPQHLIONPN PFRAAVTVFLMQEKKHEETV"
complement(1137..1141)
/gene="HCM2.0005c"
/note="possible RBS"
complement(11492..1497)
/note="possible RBS"
1565..1621
/note="repeat 1: 57 bp repeat"
complement(1647..2489)
/gene="HCM2.0006c"
CDS
complement(1647..2489)
/gene="HCM2.0006c"
/note="HCM2.0006c, possible lipoprotein, len: 280 aa; shows weak similarity to TR:AAF10061 (EMBL:AE001907), DR0482, Deinococcus radiodurans B-cell receptor associated protein-related protein (328 aa), fasta scores: E(): 4.9e-07, 23.4% identity in 265 aa overlap. Highly similar to TR:CAB55231 (EMBL:AL117211), YPWT1.49C, Yersinia pestis CO-92 putative lipoprotein from plasmid pMT1 (276 aa), fasta scores: E(): 0, 97.5% identity in 280 aa overlap and TR:O68763 (EMBL:AF074611), Y1114, Yersinia pestis K1M5 hypothetical protein from plasmid pMT1 (276 aa), fasta scores: E(): 0, 97.5% identity in 280 aa overlap. Contains a possible N-terminal signal sequence and an appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="putative lipoprotein"
/protein_id="CAD09873.1"
/db_xref="GI:16505987"
/translation="MRKGLLAVALTAICTMGLTGCDRVEPGYVIGKYNLGEDKGIGE VVGVRQWGLNTELYTPTFKOMKTYDEPTFQMSDGTAGHKIGVAYLVNRRDKVTT VFQYRKGVDDITDTRQKIADSLNRLASRMTDTTFIDGSKASLLNALKDIAEMS PVGIEVITSLSWGVKPDYPTDTVIESINAKTTLQRQOQVEVQRKAENMLREQAGG

```

Db 94243 CAACCTGGCGCCCACTAACAACCTTGAAGACTGGCCACTGATTCGAACAAC 94193

## RESULT 31

## STRMAG

LOCUS STRMAG 1555 bp DNA linear BCT 16-AUG-1994  
 DEFINITION Streptococcus dysgalactiae (mag) gene, complete cds.  
 ACCESSION L27798  
 VERSION L27798.1 GI:475117  
 KEYWORDS mag gene.  
 SOURCE Streptococcus dysgalactiae DNA.  
 ORGANISM Streptococcus dysgalactiae  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1 (bases 1 to 1555)  
 AUTHORS Jonsson,H., Frykberg,L., Rantamaeki,L. and Guss,B.  
 TITLE MAG, a novel plasma protein receptor from Streptococcus  
 dysgalactiae  
 JOURNAL J Biol Chem 270: 249-254 (1995)  
 MEDLINE 94259307  
 FEATURES

## FEATURES

Location/Qualifiers

1..1555  
 /organism="Streptococcus dysgalactiae"

/db\_xref="taxon:1334"

-35\_signal

-10\_signal 230..235

-10\_signal 249..254

-10\_signal 264..269

RBS 276..281

gene 288..1529

/gene="mag"

/gene="mag"

/gene="mag"

/function="binding of macroglobulin, albumin and IgG"

/codon\_start=1

/transl\_table=11

/protein\_id="AAA26921.1"

/db\_xref="GI:475118"

/translation="MEKEKVKYFLURKSAFLGSLVSAFLVGTAVVNEESTVSPVTV

ATDAVTTKEALAIINKLSEDLNLDIQEVLAKAGRDILASDSADTIKALLAEVTA

VTRENEKMDADVDKAIADAAAFSELKDALQKAYEDLAKLAADTDLDLVAKIIND

YTKVENAKTEDVKKIFESQNEVTRIKTERKALAKAKADAIEILKYGIGDY

YIKLNNGKTAEGVTLKDEILASKPAVIDAPELTPALTYKLVLNGTLKGETTTKA

VDLAEAKFAFYANENGVDGVVYDDATKFTFTVMVTEVPGDAPTEPKPEASIP

VLPATPIAKDDAKDKDTKKDDTKKDAKKPEAKKEAKKAATLPTTGECSNPFFTA

AALAVMAGALAVASKRKED"

288..389

/gene="sig\_peptide"

/note="putative"

390..1526

/gene="mat\_peptide"

/note="putative"

/function="binding of macroglobulin, albumin and IgG"

/product="unnamed"

BASE COUNT 582 a 242 c 305 g 426 t

ORIGIN

## Alignment Scores:

Pred. No.: 1..83 Length: 1555  
 Score: 114.50 Matches: 49  
 Percent Similarity: 34.22% Conservative: 15  
 Best Local Similarity: 26.20% Mismatches: 56  
 Query Match: 14.74% Indels: 6  
 DB: 1 Gaps: 6

US-09-847-539a-6 (1-159) x STRMAG (1-1555)

Qy 36 AlaAlaAspGluLeuLysLysGlnAlaLeuGluAsp----- 47  
 Db 903 GCTATTGAATTCGTGAAGAATACGGAATTCGCGATTACTATTAATTAATAAT 962  
 Qy 48 -----LysGluAlaThrThrAlaLeuGlu-----AlaAlaSerSerAspAla 61  
 Db 963 GGTAAACTGCAGAGGTGTGCTGCTCTTAAAGATGAATTTTAGCTTCAAAACAGCA 1022

Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 81  
 Db 1023 GTGATTGACGCACCTGAAATTACACACGCTTTGACAACTTACAACTTGTATC----- 1076  
 Qy 82 AlaAspAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101  
 Db 1077 -----AATGGTAAACATTTGAAAGCGCAACAACTACTAAAGCAGTAGACGCAAACT 1130  
 Qy 102 SerGluGluAla---GluValValGlnSerAspAlaAlaSerAspAlaThrGlu--- 119  
 Db 1131 GCAGAAAAACGCTTCAACAACTACGCTTAAAGAAACGGTGTGATGGTGTGGACTTAC 1190  
 Qy 119 ----- 119  
 Db 1191 GATGATGCCACTAAGACCTTTACTGTAAGTGAATGCTTACTGTAAGTCTGCTGTATGCA 1250  
 Qy 120 -----LysAlaAlaThr 123  
 Db 1251 CCAACTGAACCAAAAAACCAGAACGATATCCCTCTTTGTCGTTAACTCCTGCAACT 1310  
 Qy 124 ProLeuAlaLeuAspValLysLysThrLysAspThr----- 135  
 Db 1311 CCAATTGCTTAAAGATGACGCTTAAAGAAAGACGATACTAAGAAAGACGATACTAAGAAAGAA 1370  
 Qy 136 -----LysProValValLysLysGluGlnAsnValAsnValAsnThrLeuProThr 152  
 Db 1371 GATGCTAAAAACCCAGAACGCTAAGAAAGAAAGAGCTAAGAAAGCTGCAACTCTTCTCTACA 1430  
 Qy 153 ThrGlyGluSerAsnPro 159  
 Db 1431 ACTGGTGAAGGAAGCAACCCA 1451  
 RESULT 32  
 STYPPHCM2/c 106516 bp DNA linear BCT 26-OCT-2001  
 LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18  
 DEFINITION plasmid pHC2.  
 ACCESSION AL513384  
 VERSION AL513384.1 GI:16505981  
 KEYWORDS Salmonella enterica subsp. enterica serovar Typhi.  
 SOURCE Salmonella enterica subsp. enterica serovar Typhi  
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Salmonella.  
 REFERENCE 1 (bases 1 to 106516)  
 AUTHORS Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,  
 Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,  
 Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,  
 Connor,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,  
 Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,  
 Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,  
 Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,  
 Stevens,K., Whitehead,S. and Barrell,B.G.  
 Complete genome sequence of a multiple drug resistant Salmonella  
 enterica serovar Typhi CT18  
 TITLE Nature 413 (6858), 848-852 (2001)  
 JOURNAL 11677608  
 PUBMED 11677608  
 REFERENCE 2 (bases 1 to 106516)  
 AUTHORS Parkhill,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-OCT-2001) Submitted on behalf of the Salmonella  
 sequencing team, Sanger Centre, Wellcome Trust Genome Campus,  
 Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk  
 NOTES:  
 Details of S. typhi sequencing at the Sanger Centre are available  
 on the World Wide Web.  
 (URL, [http://www.sanger.ac.uk/projects/S\\_typhi/](http://www.sanger.ac.uk/projects/S_typhi/)).  
 FEATURES  
 source  
 1..106516  
 /organism="Salmonella enterica subsp. enterica serovar  
 Typhi"  
 /plasmid="pHC2"  
 /strain="CT18"

```

Db 13433 CGCATCGCTCCACATCTCGCCGACCTTTGGTTGAAGCTGTCAACGATGTTCCAGCCAGTGTT 13492
Qy 25 ---ProGluLysLeuAlaLeuArg-
|||
:::|||||
Db 13493 GAGCCTTCAGCGCTGCTTTGCGTGCCTGCGCGCTAAATTCGTTTCTGATCTTGAAGAAG 13552
Qy 32 -----AsnGluGluArgAlaIleAspGluLeuLysLysGlnAla--- 44
|||
:::|||||
Db 13553 AACATCCAAAAGACTGTGAGCATGTAAGAAGCGCCGATGAACCTGAAGAAGATGTCGGA 13612
Qy 45 -----IleGluAsp-
|||
:::|||||
Db 13613 AATCGGTCAAGACGCTGGAAGATGTTGTTTCGAAATACCAAAATCAACCTCAACCACT 13672
Qy 55 GluAlaAlaSerSerAspAla-
|||
:::|||||
Db 13673 GACGTGGCTAAGGACGAGCTCAACAACTCAAGCTACTGTAGAGCAACTTACCAAGCTG 13732
Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAlaAspAla 85
|||
:::|||||
Db 13733 GCCAGTCTAGCACAAGATGATCCCAAGTCCCAAGTCCCAAGGACATCAAGGACTCAAGACT 13792
Qy 86 AlaSer-
|||
:::|||||
Db 13793 AAGCTTAGGAATTCGTTCCAGCTCTTGAAGAGCCATTCACCAAGAGGAGCTATTTCGT 13852
Qy 102 SerGluGluAlaGluValValGlnSerAspAlaAlaSerAspAlaTrpGluLysAla 121
|||
:::|||||
Db 13853 CGTCAACAAGCGAGATT-----AATGATCGTCTCAACAATCTCGAAGGAG 13900
Qy 122 AlaThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLysLys 141
|||
:::|||||
Db 13901 CTTACCAAGGTG-----GATGAGTTCACCAAGGAGGATGCACTTCCAATTTGTT---GAC 13951
Qy 142 GluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsn 158
|||
:::|||||
Db 13952 CAATCGCGCCCACTACTACACCTTGAAGACTGCCACTGATTCGCAACAAC 14002

RESULT 30
AC006902/c
LOCUS
DEFINITION
AC006902
Caenorhabditis elegans clone Y7A11Y, *** SEQUENCING IN PROGRESS
***, 11 unordered pieces.
AC006902
AC006902.2 GI:4309900
VERSION
HTG: HTGS_PHASE1.
KEYWORDS
Caenorhabditis elegans.
SOURCE
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
Waterston,R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 96468)
Waterston,R.H.
Direct Submission
Submitted (24-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Mar 1, 1999 this sequence version replaced gi:4263446.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2040: contig of 2040 bp in length
* 2041 2055: gap of unknown length
* 2056 10960: contig of 8905 bp in length
* 10961 10975: gap of unknown length

```

```

* 10976 21821: contig of 10846 bp in length
* 21822 21836: gap of unknown length
* 21837 30764: contig of 8928 bp in length
* 30765 30779: gap of unknown length
* 30780 37500: contig of 6721 bp in length
* 37501 37515: gap of unknown length
* 37516 47190: contig of 9675 bp in length
* 47191 47205: gap of unknown length
* 47206 57255: contig of 10050 bp in length
* 57256 57270: gap of unknown length
* 57271 68424: contig of 11154 bp in length
* 68425 68439: gap of unknown length
* 68440 79779: contig of 11339 bp in length
* 79780 93036: contig of 13243 bp in length
* 93037 93051: gap of unknown length
* 93052 96468: contig of 3417 bp in length.
FEATURES
Location/Qualifiers
1..96468
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y7A11Y"
BASE COUNT 30222 a 17588 c 18100 g 30408 t 150 others
ORIGIN
Alignment Scores:
Pred. No.: 97.7 Length: 96468
Score: 116.00 Matches: 49
Percent Similarity: 39.09% Conservative: 28
Best Local Similarity: 24.87% Mismatches: 66
Query Match: 14.93% Indels: 54
DB: 2 Gaps: 9
US-09-847-539A-6 (1-159) x AC006902 (1-96468)
Qy 9 ArgIleIleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAla----- 24
||||| |||
Db 94762 CGCATCGCTCCACATCTCGCCACTTTGTTGAAGCTGTCAACGATGTTCCAGCCAGTGTT 94703
Qy 25 ---ProGluLysLeuAlaLeuArg-
|||
:::|||||
Db 94702 GAGCCTTCAGCGCTGCTTTTCGCTGATCGTCCCGCTAAATTCGTTTCTGATCTTGAAGAAG 94643
Qy 32 -----AsnGluGluArgAlaIleAspGluLeuLysLysGlnAla--- 44
|||
:::|||||
Db 94642 AACATCCAAAAGACTGTGAGCATGTAAGAAGCGCCGATGAACCTGAAGAAGATGTCGGA 94583
Qy 45 -----IleGluAsp-
|||
:::|||||
Db 94582 AATCGGTCTAAGAAGCTGGAAGATGTTGTTTCGAAATACCAAAATCAACCTCAACCACTC 94523
Qy 55 GluAlaAlaSerSerAspAla-
|||
:::|||||
Db 94522 GAGCTGCTAAGGACGAGCTTAACAACTCAAGGCTACTGTAGAGCAACTTACCAAGCTG 94463
Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAlaAspAla 85
|||
:::|||||
Db 94462 GCCAGTCTAGCACAAGATGATCCCAAGTCCCAAGGACATCAAGGACTCAAGACT 94403
Qy 86 AlaSer-
|||
:::|||||
Db 94402 AAGCTAAGGAATTCGTTTCAGGCTCTTGAAGAGCCATTCACCAAGAGGAGCGCTATTTCGT 94343
Qy 102 SerGluGluAlaGluValValGlnSerAspAlaAlaSerAspAlaTrpGluLysAla 121
|||
:::|||||
Db 94342 CGTGAACAGCCGAGATT-----AATGATCGTCTCAACAATCTCGAAGGAGAG 94295
Qy 122 AlaThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLysLys 141
|||
:::|||||
Db 94294 CTTACCAAGGTG-----GATGAGTTCACCAAGGAGGATGCACTTCCAATTTGTT---GAC 94244
Qy 142 GluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsn 158
|||
:::|||||

```









```
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99069613
2 (bases 1 to 54126)
Gattung,S., Goela,D. and Brody,M.
The sequence of C. elegans cosmid 2K973
Unpublished
3 (bases 1 to 54126)
Waterston,R.
Direct Submission
4 (bases 1 to 54126)
Unpublished
Waterston,R.H.
Direct Submission
Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 54126)
Waterston,R.H.
Direct Submission
Submitted (01-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 54126)
Waterston,R.
Direct Submission
Submitted (08-MAR-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 54126)
Waterston,R.
Direct Submission
Submitted (28-MAR-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
8 (bases 1 to 54126)
Waterston,R.
Direct Submission
Submitted (03-OCT-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
On Oct 5, 2001 this sequence version replaced gi:4263104.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: rwenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one ml3 subclone.

NOTES:
Coding sequences below are predicted from computer analysis, using
the program GeneFINDER(P. Green and L. Hillier, ms in preparation).
FEATURES
Source
1..54126
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="1"
/clone="2K973"
11..25772
/gene="ZK973.6"

join(11..103,170..209,1232..1563,1609..1736,1784..1940,
1993..2127,2196..3029,3082..3300,3352..4719,4767..5015,
5055..9450,9629..12490,12669..15377,15556..18264,
18443..21151,21330..24036,24217..24654,24700..24843,
24890..25021,25098..25586,25635..25772)
/gene="ZK973.6"
/note="contains similarity to myosin-like proteins; coded
for by the following C. elegans cDNAs: cm5b12, cm10f6,
cm15g2, yk2f4.3, yk2f4.5, yk7e10.3, yk7e10.5, yk19b4.3,
yk19b4.5, yk32c4.3, yk32c4.5, yk36b3.3, yk36b3.5,
CEES066F, CEMSA37F, CEMSH10F, yk25a11.5, yk8f4.5,
yk752h02.5, yk871h03.5, yk54d5.5, yk855c06.5, yk44g3.5,
yk539e4.5, yk542g12.5, yk369e5.5, yk881h02.3, yk539e4.3"
/codon_start=1
/product="Hypothetical protein ZK973.6"
/protein_id="AAF40010.2"
/db_xref="GI:15967153"
/translation="MELFRRLLNFCDAVKILRSQNTKWNKIKISQCCRRGGGEYIIGS
ECGFAEAFCKGNLFTSLVSCFQIYLCLFSQOKLVISRVLFSLKLTTCGCPLLF
SCTCRSPSLTLMSLFORSKIYHTRFLFSORRGLODMANRTRTHLKHESKIDFT
LIVRLDEWTDONOLRAELASIHGSPFQLEDFRLNEKLMILHDKDKSLMLVHRRH
YLTAELNLTDSKTDLAQRATSSSDPDDILRATDEYTKALDIKEELRRLGELAMNI
TDLVVVALISFRNRQSGEPEVEEURLARELIARTEPRDHPDPAIVADILRM
KDEKKEKTIDEQTATLTDEQRASFAPLIEDYRRADRHRIYFEHLVLMYLSLRL
QFDELEIGTITQTSRADDLRMNSTENKWKTDLANIERQVQDPTKKALSADL
HRRKDSMEARINKYLTHSAKINAKLTQFKWLNAEEDIEQTFQEPDERSYRFGSL
HEVALAKRLVAKLERLVANKEEVLCERYHTMHKLTPTQTAGVPLHVSYNLDR
NGPQSQISVSSIASSELERPESVMSLTSSIGVTPADVAELSPPEAKTNKLQKHII
EDSLGKPKPIDTVRRQVQLEKYNRGAELIQLQSTSNIEDAEKGGUKRFLMLNN
YDDLRSSTENIRDNELTAKNBEILAELSNAQTQNSPLEDDIDAEELRLQMLND
LIVKMNPKRYVECELIDSSREGSPQRRRRKKVMVGNVNTITTIHVVEERLEAS
DISRNPQVKLVAKENLRDLTITVTPHPPSIMPCLPTNRNDELVEVKLAAEIDR
AIDPASMVEDAPTDADKSAIHLDDQKVTLNHLVHLEGIEPKNQDRTDADIA
SSVGEKGLNLSAVEEYVEVLASTNPVKEDQLOHQIOEQVTTAEQNSQDDESRQ
PPVLSDERIELKTDPSAIDKFEVRSDEDPVKIIDLQLOQTAVNEASPLACETDDV
DALQVADSKLTQDRTIRKIHAILDITDDPVQKPAIIEISLDRQKIDNNRNNRQI
DNLNNTQTPVVPVKESTPLNEEDAVRQASTVVSDELCTNKILLSARQTLANRQKQ
VDSVEAVVSNCPETIETVPLEEYTLVDNIEKRLANEIEVPTNDPQROLOOQLOD
VILCEEVVNCNTEKUESKLLEKARPLDQIDGNVEKLSRQSPDSDAIDALS
NVHQYNATINSIDDKIDELKNPDETSADQILSELHVIEMPAVTIDLSMLNAIEE
GLSTPLHQAENVOAKIDELQKKEVADQTEQLISDLNAGFMDPAITLIDLNLKLEE
GLIPLVEDSRIRAKIVDLRRKKEDQDAALLOELSVISDMPLSTLDLNNMQIED
NLNSLPAEEDIRKINLRRRKOESDAALLOELSVISDMPLSTLDLNNMQIED
NLNSLPAEEDIRKINLRRRKOESDAALLOELSVISDMPLSTLDLNNMQIED
NLNSLPAEEDIRKINLRRRKOESDAALLOELSVISDMPLSTLDLNNMQIED
NLNSLPAEEDIRKINLRRRKOESDAALLOELSVISDMPLSTLDLNNMQIED
GLSTLLSEDRSKIQQAIDSRLKKSDSLAQHALSALSVQSKLPSVSNLBELKLEE
TLSTVPLVEDSRIRAKIVDLRRKKEDQDAALLOELSVISDMPLSTLDLNNMQIED
QILQMPVOYQPSVKETLDLQKAEEDDKLAGVYDELEKIAKLPAQYDNKLAKIDE
KINSIPKQIATETHRKVEDIKVTRADIVQIDVLDKLPKADIDEHLNLSIEKLTPI
SDSDLOQIAGIKLRDKRKOANIDCKKTLNELAEIOKMPADSLNEHLNLLATESDKF
GSEISDKTMOEDIVLRKQNNHEVARLNAESVLOQLDKISEEPHLSLTERLAPLQV
IDTVPACFVDKIRNKINVEQKHQENODEKELKIVAKVONIGKTSIDVAVDFE
BEIERINGSLEAFEPPLAKIOELREKARVGDGARSAAHQIIVALEKREADVTAKE
SAKKKKDKKSPQEMIDELSAVNAKALPKIEEAAKNENLPADPKPAEQVLSNL
EAFVKDVEQVSEKODELKNANADAKRLGDALDDAEKTVVSPSPALSPFKDRIA
PHLATLVAVNDVPASVPSVALDRAAKFVSDLEKNIQKTDGDEKADKNDVGN
AVKVEDVSVYQVQPOPLDVAKDDANKATVEOLTKLAEESDKIDQVAKDKIDSK
TKAKELQALEKALPOEDAIREQAEINDRNLNLEKLTVDPEKPEKALPVLQOLA
NTNLTKATDSNNEKAVAPSSILSHDDLVGGLPEKVFQLOHAIIDKKQALNKAAYNE
LAPKLVLSVQQLQSVQPEVPASLDEQKOLLEDVENQHNLENLANLENPDTELELR
OKSOWDSRLKDLKLGSAVGDKLAALAFNAARKNAEDALLDITREDGDNNKSPD
ELIDDLAKKEETVAKLIDTVGVKPDDELDKERAENLDLARLATADADNKRKRAELE
QAVKAKADEKSLHDSVDRIVSRLVPLVRESDELHNAEAVPTQVAPKAEELKEVEEA
KAVTANAPSSDAHVQQLQEQAVATAETLIPDLLEEARLWNEFLARNDELDALIEQLQV
LDVLAQPKRAEAAEQVQENLRNNSQOLSOLDNKNIANLRISELLDPESADQVRF
FDVDAEQTRHQYDDVLDNVAELEDETLKQSAQVANEIDDIKMDIDTDEPSRILD
TIAKSDIPALKAQINRIVRNADASKRHVTTPDKIAEDLONLAKIQTDLDAIKT
SDEHDKQILSLSKNISOFOIPLDLOKSDOLKTAKEKTNLSLPEAEELPLAKIOE
LREAKRVGDEARSAAHQIIVALEKREADVTAKEKAKKKKKKSPQEMIDELSAKVV
ZAKALIPKIEEAAKNENLPADPKPAEQVLSNLEAFVYKDVETQVSEKQDELDKNNAN
DAIKRLGDALDDAEKTVVSPVSPALSEFKDRIAPHLATLVAVNDVPASVPSVALR
DRAAKFVSDLEKNIQKTDGDEKRADELKNDGNAVKNVEDVSVYQVQPOPLDVAKDD
```







```
/transl_table=11
/evidence-not_experimental
/product="putative polyprotein"
/protein_id="BAB34846.1"
/db_xref="GI:13360887"
/translation="MKTRHVVQDDITKLVADVIVNAANSLMGGGGVDAIIRAAGP
ALLDACLVKQQQDCPTGHAVITLAGDLPAKAVVHTVGPVWRKGGEQNEOLLQDAYL
NSLRVAANSSTVAFPAISITGVYGPRAAAAEIAVKTVEFITHRALPQVYFVCYD
EENAHLYERLLTQQGDE"
549..2030
/gene="ECs1424"
549..2030
/gene="ECs1424"
/gene="ECs1424"
/note="similar to YMDC_ECOLI gi|1787284 percent identity
99 in 493 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative synthase"
/protein_id="BAB34847.1"
/db_xref="GI:13360888"
/translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSQHPOGCGLPFL
EKSIDAAARYLAEMSEHTLDVQYIWDQDMSGRLLFSALLAAAKRGVVRLLLDN
NTPGDDILLRDSHPRIEVRLEPNFSLRULPUGYITDFSRLMRRHNSKSFVTDGV
TLVGRNIGDAYFGAGEEPLFSDLOVMAIGPVVEDVADDFAFYWKSVSPLOQVLDV
PEGEMADRIELPASWHNDAMTHRYLRKXESSPFINHLDVGTPLIMAKTLLSDPAK
GEGKAKRSLIPORLFDITMGSPSERIDIISSEYFVPTAGVAOLLRVMRKGVKIALTN
SLAANDVAVVAGVWRKLLRYGVVELYELKPTREOSSTLHDKGITGNSGASIAHKT
FSIDGKTVFISGFNFDPSTLLNTEMGFVIESETLAOLIDKRFLOSQYDAWQLRLDR
WGRINWDRHAKKEIVLKKEPATSFWRVVRVRLASILPVEWLL"
complement(2038..3195)
/gene="ECs1425"
complement(2038..3195)
/gene="ECs1425"
/note="similar to MDOC_ECOLI gi|3025226sp|p75920 percent
identity 99 in 385 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="glucans biosynthesis protein"
/protein_id="BAB34848.1"
/db_xref="GI:13360889"
/translation="MNPVPAQREYFLDSIRAWMLLGLIPFHISLIYSKHTWHVNSAEP
SLWLTNDFTHSRMQVFFVIGSYFIMFLRYPLRKKMKWVRVERVGIPMLTAIPLL
TLPOFIMQVYKGAESWPGLSLDKYNTLAWELISHLWLLVLMVTTLCVWTFKRI
RNNELESDKTNKFSWKLSVIRICLIGIGYAVIRTIIVYPPITLSNGTFNFIWQTL
FYLPFFILGALAFIPEHLKALFTTPSRGCTLAALAFVALLNORSGSDAWMYETES
VITVGLWMVNVVFSFGHRLNLFQASRVTFVNASLFIYLVHHPILFTFFGAYITPHI
TSNWGLFGLGIFVVGITAILYEIHLRIPLILKFLFSKPVVYKRENDKAPAR"
3588..5123
/gene="ECs1426"
3588..5123
/gene="ECs1426"
/note="similar to MDQG_ECOLI gi|1787286 percent identity
99 in 511 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="periplasmic glucans biosynthesis protein"
/protein_id="BAB34849.1"
/db_xref="GI:13360890"
/translation="MMKMRWLSAAVMLTYTSSWAFSIDDVAQQAQSLAGKGYEAPK
SNLPSVFRDMKYAQTOQFNHDKYWNKLTPEKLEFYHOGYFDTVPK INEYATATA
VKRIKSPDYFTFDGVDHDKDVTVDLGFAGFKVLYPINSKDKNDIYVSLMGASVRYI
CAGOVYGLSAGLADIALPSGEEFPEKPIERKPTDKRLTIYALLSPRATGAY
KRYVMPGRSDTVVDQSKIIYLRDKVGLKVAPLTSMFLFQPNQPSANNRPELHDSNG
LSIHAGNGEWTWRPNLNPKLHVASFSNENPQGLLQGRDFFSRFEDLDDYDLRPS
AMWTPKGMWKGSELYEIVPTNDENDNIYAVTYPDQLPEPKENMEKFTITFDRDED
KLHAPDNAWQVTRRSYGDVKOSNLIRPODGTIAFVVDFTGAEMKKLPEDTPVTAOTS
IGDNGEIVESTVRYNPVTKGWRVLRVYKVDKAKKTTMRAALVNADQTLSETWSYQLP
ANE"
5086..7659
/gene="ECs1427"
```

```
CDS
5086..7659
/gene="ECs1427"
/note="similar to membrane glycosyltransferase MDOH_ECOLI
gi|1787287 percent identity 100 in 847 aa but differs at
N-ter (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="membrane glycosyltransferase"
/protein_id="BAB34850.1"
/db_xref="GI:13360891"
/translation="MRPGATSYLPMKTTETIDAMPIAASEAKALPKTDIRAVHQALD
AERTWAREDSDPGSVKARLEQAWPDSLADGLIKDDGRDQLKAMPEAKRSMEFPD
PRTNPNVRFWRDLGRDVTPLYRLARTKEEQSKWRTVGTIRYVILLILLAQTV
VATVWKTILPYQWALINPDMVMGQDLVSMFQMLPYMLQGTGILILFAVLPQVWSAG
FWTALMGFLOLLIGRDKYSISASTVGEPLNPEHRTALIMPICNEOVNRFVAGLRATW
ESVKATGNAKHFDVYILSDSYNPDICVAEQKAMNELIAEVGEGOIIFYRRRRVRKRK
SGNIDPCRRWGSQYSYMWVLDADSVMTGCLGLVRLMEANWAGIIQSSPKASMD
TLYARCOQFATRYVGLPTAGLHFWQLGESHYWGHNATIRVKPFIEHCALAPLPGEGS
FAGSILSHDFEAAALRRAGWGVIAVDLPGSYEELPPNLLDELKRRWRCHGNLMNF
RLFLVKGMPHVRHAFVLTGVMSYLSAPLWFMFLASTALQVHALTPQVYFLOPRLF
PVWPPRPELAIALFASTVMYLLFLPKLLSILLWCKGTKEYGGFWRTLSLLEVLFS
VILAPVRLMHTYFVVSFAFLGNEVWNSPORDDDSTWGEAFKHKCSOLLGLWANG
MAMDURFLFWLAPTVFSLSPFSVSISSRAIVGLTKRWKLFLLIPEEISPPQVLYVD
TDRFLENRORSDDGGMFAVFNPSFNALATAMATARKASKVLEIARHRHVEQALNE
TPKELNDRDLVLLSDPVTWARLHFRVWNSPERYSWVSYIEGILKLNPLALRKPDAAS
Q"
7832..8059
/gene="ECs1428"
7832..8059
/gene="ECs1428"
/note="identical to YCEK_ECOLI gi|1787288 (Conserved in
E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34851.1"
/db_xref="GI:13360892"
/translation="MRLIVYSIMVTLSSGCSIIISRTIPQGGHGNQYYPGVQMDVRDS
AWRYVTLDLPFSLVFTLLLPIDIHGPEY"
complement(8060..8437)
/gene="ECs1429"
complement(8060..8437)
/gene="ECs1429"
/note="similar to MSYB_ECOLI gi|1787289 percent identity
99 in 125 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="acidic protein MsyB"
/protein_id="BAB34852.1"
/db_xref="GI:13360893"
/translation="MMTYATLEAIDAAREEFLADNPIDAEDANVQQPNAQYVILQ
DGDIMQVEFFADEGEGEGLPMLSGEAGSVFDGDDYDEIIRHWEQENLTLEHWDG
```

## Alignment Scores:

Pred. No.:	124	Length:	222605
Score:	119.50	Matches:	45
Percent Similarity:	44.59%	Conservative:	21
Best Local Similarity:	30.41%	Mismatches:	69
Query Match:	15.38%	Indels:	13
DB:	1	Gaps:	4

US-09-847-539a-6 (1-159) x AP002555 (1-222605)

Qy	15	GlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlu-----LysLeuAlaLeu	30
Db	184876	GGTACGCTGAATGATTTCTTGGTCCCATGACTGAGGATGATGTCCGTCGGAGGCACTG	184935
Qy	31	ArgAsnGluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAla	50
Db	184936	CGCCGTTTGTGACTGATGGTGAAGAGGTGGCGGTAAACGCGTCCGCGGTGGCACAGAAC	184995

```
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="Residues 1 to 191 of 193 are 94.76 pct identical to
residues 1 to 191 of 191 from GenPept 118 :
gi|2367120|gb|AAC74619.1| (AE000252) orf, hypothetical
protein [Escherichia coli]"
/codon_start=1
/transl_table=11
/product="putative tail fiber protein of prophage CP-933X"
/protein_id="AAG56009.1"
/db_xref="GI:12514849"
/translation="MAFRMSEQPRTKIYNLLTGTNEFIGEDAVIPHTGLPANSTY
IAPPDPAFVAVFNSDESSWHLVEDHGKTVYDVASGDALFISELGSIPENVWLSLP
EGEFQKNGTAWKDAEAEKLFRIREABETKNSLMQVASEHIAPLQDAVDLEIATEEE
ASLLCAWKRYRLLNRVDYSTADIEWPVPEV"
gene
complement(7941..8831)
CDS
complement(7941..8831)
/gene="21921"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="Residues 1 to 289 of 296 are 84.42 pct identical to
residues 1 to 289 of 292 from GenPept 118 :
gi|7162108|emb|CAB76676.1| (Y19180) putative catalase
[Salmonella typhimurium]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933X"
/protein_id="AAG56010.1"
```

## Alignment Scores:

```
Pred. No.: 5,34 Length: 10432
Score: 119.50 Matches: 45
Percent Similarity: 44.59% Conservative: 21
Best Local Similarity: 30.41% Mismatches: 69
Query Match: 15.38% Indels: 13
DB: 1 Gaps: 4
```

US-09-847-539a-6 (1-159) x AE005333 (1-10432)

```
QY 15 GlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlu-----LysLeuAlaLeu 30
Db 4574 GCTACGCTGAATGATTTCTTGCTGCCACTGAGGATGATCGTCGGAGGCACCTG 4633
QY 31 ArgAsnGluGluArgAlaIleAspGluLeuLysLysGluAlaIleGluAspLysGluAla 50
Db 4634 CGCCGTTTTCGAGTGTATGGGAAGTGGCGGTAACCGCTCCCGGTGGCAGCAGAAC 4693
QY 51 ThrThrAlaIleGluAlaAlaSerSerAsp-----AlaLeuGluAlaLeuAla 66
Db 4694 ACGGACGCCCGGGAAGTACGCCACGATGCCAGCACATCAGCCGTCGAGGCGGCAACC 4753
QY 67 AspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLysAlaAspAsnAlaAla 86
Db 4754 CATGCGACTGATGCTCGGACTCAGCAGCGAGCGGACGACGTCAGCGGACAGGCGCG 4813
QY 87 SerAspAlaLeuGluAlaLeuAlaAspGln---ThrAspAlaLeuGlnSerGluGluAla 105
Db 4814 TCGTCGGCTCAGTCACGCTCTCCACGCGGAGGAACGATCGCGCAGTGTCAACAAGTACTGAAGCA 4873
QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTirpGluLysAlaAlaThrProIle 125
Db 4874 -----TCAAAAGTCTGCGCTGCGAGATCTCTCAAAAGCGGCGGCTACC 4921
QY 126 AlaLeuAspValLysThrLysAspThrLysProValLysLysGluAlaArgGln 145
Db 4922 AGTCCGCGTCCGCGGAAACGTCAGAAACGAATGCGGCGAGTGTCAACAAGTACTGACGCCGC 4981
QY 146 AsnValAsnThrLeuProThrThr 153
Db 4982 ACTTCTGCATCCACCGGACCGAGC 5005
RESULT 23
```

```
AP002555 222605 bp DNA linear BCT 07-MAR-2001
Escherichia coli O157:H7 DNA, complete genome, section 6/20.
AP002555 BA000007
AP002555.1 GI:13360886
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (sites)
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C., and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 222605)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan [E-mail:ken@gen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047]
genome project.
Location/Qualifiers
1..222605
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RMD 0509952"
/db_xref="taxon:83334"
74..607
/gene="ECS1423"
74..607
/gene="ECS1423"
/note="similar to B1045_ECOLI gi|1787283 percent identity
100 in 177 aa (Conserved in E.coli K-12)"
/codon_start=1
```

Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouasis,K.,  
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,  
 Welch,R.A. and Blattner,F.R.  
 Direct Submission  
 Submitted (22-Oct-2000) Laboratory of Genetics, University of  
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 Location/Qualifiers  
 source  
 1. .10432  
 /organism="Escherichia coli O157:H7 EDL933"  
 /strain="EDL933"  
 /serotype="O157:H7"  
 /db\_xref="taxon:155864"  
 /note="enterohemorrhagic"  
 <1. .>10432  
 /note="O-island #52; Region of the EDL933 chromosome not  
 homologous to E. coli K-12 MG1655"  
 130. .1296  
 /gene="21915"  
 130. .1296  
 /gene="21915"  
 /function="putative structure; Structural component (Phase  
 or Prophage Related)"  
 /note="Residues 8 to 366 of 388 are 81.05 pct identical to  
 residues 11 to 369 of 1132 from GenPept 118 :  
 gi|215125|gb|AA96553.1| (J02459) J (tail:host  
 specificity:1132) [bacteriophage lambda]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative tail protein (partial) of prophage  
 CP-933X"  
 /protein\_id="AAG56004.1"  
 /db\_xref="GI:12514844"  
 /translation="MAARTREVEAKDNLKSTQMMSVIDAIGEGPIEGPVKGLQSLIV  
 NKTPLTDGNGPIHGVTAVMRAGEQETPPGESSGAETALGVEVTKAKPTVTIT  
 SANIDRLVTFVGSLLFTSKGDRNHSSVRLLIQORGNWVTEKDYITNGKTTTSQF  
 LASVILDLNPPRFNIRMVRETDSTTQOLQRLTWSSYTEIIDVKQCPNPAIYGLQ  
 VDAPFGQOQWTVNHIIGRIITQVFNSTDPKRTYSGWDGSLKPAYSNPAWCLWDM  
 LTRPIGMRKRGUAGADVKWALYAIAQYCDQVDPGFGTEPRMTFNAYLSOORKWD  
 VLSDPSCMRGKMPVWNGQTLTFVDRPDSVVMVPTPAVMMWMTAWGFATASAP"  
 1245. .3593  
 /gene="21916"  
 1245. .3593  
 /gene="21916"  
 /function="putative structure; Structural component (Phase  
 or Prophage Related)"  
 /note="Residues 1 to 781 of 782 are 62.37 pct identical  
 to residues 376 to 1131 of 1132 from GenPept 118 :  
 gi|215125|gb|AA96553.1| (J02459) J (tail:host  
 specificity:1132) [bacteriophage lambda]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative tail protein (partial) of prophage  
 CP-933X"  
 /protein\_id="AAG56005.1"  
 /db\_xref="GI:12514845"  
 /translation="MVVDNGVGFYRFSALKDRHTAVEVNTDPQNGWOTSTELVED  
 PPAILRYNGLKMDAGCTSGQARHAGLWIKTLLLETOTVDFTLGSGQLRHTPGD  
 IIEICNDYAGMTGGRVLSIDAASKTLLDREVTLPETGAATVNLINGSGKPVSDVI  
 TAAHPARDIOVSTLPDGVETGVWGLSLPCLRLFCVSVRENTDGTFAITAVQHP  
 EKEAIVDNGAREPQSGSLNSVIPAVQHLTVESVADQYLQAKWDTTPRVKGVRF  
 SLRLTSGKTDARLTALTADTHERFSGCLPIGEYTLTVRAINSGQDGEPAITTFRI  
 AAPAPSRITLPGFYOTATPHLAVDPTQFVEFWFSEKRIADIRQVETAAARYLGS  
 LWIAASRIELKPGHYFYFIRSVNTVGSFSAVEAGRASDAEGYDFDFKGIKGLH  
 AQELWQIDQINGLAPDLAEIRTSITNVSNEITOTVNNKLENSAATQIOIKVOVDTN  
 NNNWAVLQOQMDGRGLYIAGIGAGIENTPAGMQSOVLLAADRLLAMINPANGNTKPM  
 FVQGDQIPMNEVFLKYITAPITSCGNPPAPFSLTPDGLRTAKNADISGNVANGSTL  
 NNTTINENRVLGKLSAQOIGDLVTKVGFAPRDSRAPRPSGTITIVRYVDQFPD  
 RQIVAPAVSGAKHEKHEKTDIYSSCRILVRKNGAEIYNRTALDNTLTIYSGVIDMPAG  
 HGHMTLEFSVANLVNNNYPTASISDLLLVVMKKATAGITIS"  
 3661. .4260  
 /gene="21917"  
 3661. .4260  
 /gene="21917"

/function="putative membrane; Other or unknown (Phage or  
 Prophage Related)"  
 /note="Residues 1 to 199 of 199 are 75.87 pct identical to  
 residues 1 to 199 of 199 from GenPept 118 :  
 gi|17532789|gb|AA963231.1|AF151091\_2 (AF151091) Lom  
 [prophage P-EiBa]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative outer membrane protein of prophage  
 CP-933X"  
 /protein\_id="AAG56006.1"  
 /db\_xref="GI:12514846"  
 /translation="MRKLCVILSAVVQVAAATPASAEHOSHTLSAGYLHASTNVPQ  
 SDDLNGINVKYEFMDALGLITSFYSAEDEQKTRYSDTRHEDSVRNRFVSYMAG  
 PSVRNINVSAYAMAGVAYSRYSTSGYDLRVTDNKGKTHDVLTGSDGDRHSNTSLAW  
 GAGVQNPXESVAIDIAIEGSGCDWRDTDGFIVGVGYK"  
 4319. .7240  
 /gene="21918"  
 4319. .7240  
 /gene="21918"  
 /function="putative membrane; Not classified (Phage or  
 Prophage Related)"  
 /note="Residues 1 to 498 of 973 are 95.78 pct identical to  
 residues 1 to 498 of 1122 from GenPept 118 :  
 gi|17876361|gb|AAC74434.1| (AE000234) putative membrane  
 protein [Escherichia coli]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative membrane protein of prophage CP-933X"  
 /protein\_id="AAG56007.1"  
 /db\_xref="GI:12514847"  
 /translation="MNAVKISGVKLGDKGTPVENCTIQLKARRNSATVVNTVASEN  
 PDEAGYGMDEVGYGOYSVILLVEGPPSHAGTITVYESQPGTLNDFLCAMTEDDVRP  
 EALREFELMVEVARNSAVQNTAAKKSADSTVTSAREATHATDAADSAARAATS  
 AGQAASSQSSASSAGTASTKATEAKSNAARAEKSAATSAAGAAKTSNTAAVVSQ  
 SAITSATATTKASEASARDASAKSAKSESSTSAASSASSASTAAGNSAKAA  
 RTSETNKKSSSTAEQSAASAKSATAALSASASTSAGQASASTAAGSAESAAS  
 SASATTKAGATEQASAAASASAAKTSATKATEAAGSATAAQAQSKSTAESAAKRA  
 SSASAKDEATROASAAKSSATSTKATEAAGSATAAQAQSKSTAESAAKRA  
 AEDIASVALEADASTTKGIVOLSSATNSTSESAAITPKAVKAAVLANGKYTAQDAT  
 TAQKGIVOLSNATNSTSEMLAATPKSVKAAVLANGKYTAQDATQKGIVOLSSATN  
 SASETLAATPKAVKAANDNANGRVPSARKVNGKALSSDITLTPDKIGLINTSTMSFSG  
 SAGFPKLAFTVMPQASSVSITLIGGAFNGVSPQQAQISELVLRAGNPNRGITGAL  
 WORTSTFNFAWNTSGDTYDIYVAIGNYATGVNIOWDYTSNASVTIHTSPAYGANK  
 PEGITDGVVSYLYTPSEQFYPPGAPIPWPSDTVPVGYALMQCQTFDKTSAYPKLAAAYP  
 SGVIFDMGWTIKGPKASGRAVLSQEDGIKSHTHSASASTDLGCTKTTSSTFDTGTS  
 TNNTCAHPSVSGTAAAGNHTHSVTGASAVSQMSQSGSVHKVVSAAVNTSAGAHT  
 HSVSGTAAAGAHHTVIGIANTHSVAIGSHGHTTITVNAAGNAENTVNAFNTVRL  
 A"  
 complement(4822. .5571)  
 /gene="21919"  
 complement(4822. .5571)  
 /gene="21919"  
 /function="orf; Other or unknown (Phage or Prophage  
 Related)"  
 /note="Residues 45 to 248 of 249 are 64.21 pct identical  
 to residues 5 to 205 of 206 from GenPept 118 :  
 gi|508991|gb|AA96556.1| (J02459) orf206 [bacteriophage  
 lambda]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown protein encoded by prophage CP-933X"  
 /protein\_id="AAG56008.1"  
 /db\_xref="GI:12514848"  
 /translation="MDAVALLFAADACLVASSFEADADDAAELADDAARAAVFE  
 ALVSDAFVSDVFAAEALAAASVAFVAVVEADDAALASDPAPPAVAAEAC  
 PALVDAALDAARAAVFEPAALCAVSEDLAFVSDVLAFAFPAVAAEALAA  
 LEALVSDFAASFEADASLAEERASDAFVAVDAEVAADCCDTAAAFVSDVFAAPALV  
 AAALFEDSAAALFDASVAFVDAVPALEAD"  
 7240. .7821  
 /gene="21920"  
 7240. .7821  
 /gene="21920"







QY 114 AlaSerAspAlaTrpGluLysAlaIaIaThrProIleAlaLeuAspValLysLysThrLys 133  
 Db 2456 CCCGAGGAGAGGTGAAGAGAGTCAGAGCCCTGTTGCACCTGAAGCTAAGAAAGCCCA 2515  
 QY 134 AspThrLysProValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThr 153  
 Db 2516 GAGCAAAAGCAGCTCCCAAAAGAGAGCACCAGCAAGGTGGAGTCA---CCCACGTCT 2572  
 QY 154 GlyGluGluSerAsnPro 159  
 Db 2573 GAGCCAGAGATGAACCA 2590  
 RESULT 21  
 LOCUS AE006458/c  
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 2 of 3, section 52 of 55  
 ACCESSION AE006458 AE005174  
 VERSION AE006458.1 GI:13259569  
 KEYWORDS  
 SOURCE Escherichia coli O157:H7 EDL933.  
 ORGANISM Escherichia coli O157:H7 EDL933  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.  
 REFERENCE 1 (bases 1 to 9331)  
 AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.  
 TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7  
 JOURNAL Nature 409 (6819), 529-533 (2001)  
 MEDLINE 21074935  
 PUBMED 11206551  
 REFERENCE 2 (bases 1 to 9331)  
 AUTHORS Plunkett,G. III.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-FEB-2001) Laboratory of Genetics, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706-1577, USA  
 FEATURES  
 source  
 1..9331  
 /organism="Escherichia coli O157:H7 EDL933"  
 /strain="EDL933"  
 /serotype="O157:H7"  
 /db\_xref="taxon:155864"  
 /note="enterohemorrhagic"  
 complement(201..731)  
 /gene="Z6025"  
 complement(201..731)  
 /gene="Z6025"  
 /function="orf; Other or unknown (Phage or Prophage Related)."  
 /note="Residues 53 to 176 of 176 are 37.50 pct identical to residues 13 to 140 of 140 from Genpept121: dbj|BA36750.1| (AB016764) ORF4 [Escherichia coli]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown protein encoded by cryptic prophage Cp-933p"  
 /protein\_id="AAK16941.1"  
 /db\_xref="GI:13259570"  
 /translation="MNVLRQAQVSSGRGEPTLGNVTSVIFNETDGRFLSSGSSGGLL TELFLYGNNGPEALRDLRLSMLSDSGAQSQESIQDKISOCKFPVSSGNFQCPEESI QCPITLERPEGVFKNSDSSAVCCLEDFDAFSLASBSYHPLRTREPTASMIISPD KCVDPKIGNFTIKDS"  
 complement(888..1157)  
 /gene="Z6026"  
 complement(888..1157)  
 /gene="Z6026"  
 /function="orf; Other or unknown (Phage or Prophage Related)."  
 /note="Residues 1 to 89 of 89 are 92.13 pct identical to

residues 41 to 129 of 129 from Genpept121: qbi|AAD25465.1|AF125520\_60 (AF125520) hypothetical protein [Bacteriophage 933W]".  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown protein encoded by cryptic prophage Cp-933p"  
 /protein\_id="AAK16942.1"  
 /db\_xref="GI:13259571"  
 /translation="MNILKLMORLCGHGKHDDREDGELLTAQLRLGPADLLESDENG IIPQDRVITQVVLADKKQIQCVVRPLQILRADCTWENIGMK"  
 complement(1159..2472)  
 /gene="Z6027"  
 complement(1159..2472)  
 /gene="Z6027"  
 /function="putative structure; Structural component (Phage or Prophage Related)."  
 /note="Residues 1 to 437 of 437 are 94.96 pct identical to residues 1 to 437 of 437 from Genpept121: dbj|BAB19560.1| (AP000400) tail fiber protein [Escherichia coli O157:H7]".  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative tail fiber protein of cryptic prophage Cp-933p"  
 /protein\_id="AAK16943.1"  
 /db\_xref="GI:13259572"  
 /translation="MTVKISGLKDGTKPVQNCITVLKARRTSSTVVVNTVASENPD EAGYSMDVEYGOYVILLVEGPPSHAGTITVYEDSQPTLNDFLGMSDDVVRPEA LRRELWVEEAARHAERAKKNAGEATISARNAGISASQAERNAANADTSAGDASESAR QAESAARAKQSEASASSASAAKASQSDAELSKRTAESAGNAARDITTA AEKARSAESAQSAESAQRIAAEEAVNRITVYVPPGPKGEPGAPGQPKGDKGERGD TPAGATGERGPADGAPGQPKGDKGERGETGLTGNAGPQPKGDTGAAGPAGPQ PKGETGAAGPVGATPGQPKGDKGETQIRFLGPASITETNSHGWFPPTDGLITGL TFLAPKDTTRVQGFQHLQVRFGDPQDVKGLDVGSDTGTCE"  
 complement(2624..3223)  
 /gene="lomp"  
 /note="Z6028"  
 complement(2624..3223)  
 /gene="lomp"  
 /function="putative membrane; Other or unknown (Phage or Prophage Related)."  
 /note="Residues 1 to 199 of 199 are 96.48 pct identical to residues 1 to 199 of 199 from Genpept121: dbj|BAB19561.1| (AP000400) outer membrane protein Lom precursor [Escherichia coli O157:H7]".  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative Lom-like outer membrane protein of cryptic prophage Cp-933p"  
 /protein\_id="AAK16944.1"  
 /db\_xref="GI:13259573"  
 /translation="MRKLCVILSAVYVQVAAATPASAAEHQSTLSAGYLHASTNVPG SDDLGNLVRYEEMDALGLITSFYSANAEDEOKTRYSDTRHEDSVRNRFWSVMAG PSVRNWFSAAYAGVAYSRSVTFYGDYLRVTNKNKTHDVLGSDSGRHSNTSLAW GAGVQNPTESTVIDIAYEGSGSDWRSDAFIVGIVGRF"  
 complement(3291..5639)  
 /gene="Z6029"  
 complement(3291..5639)  
 /gene="Z6029"  
 /function="putative structure; Structural component (Phage or Prophage Related)."  
 /note="Residues 1 to 782 of 782 are 99.87 pct identical to residues 25 to 806 of 806 from Genpept121: dbj|BAB19562.1| (AP000400) host specificity protein J [Escherichia coli O157:H7]".  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative tail component of cryptic prophage Cp-933p"  
 /protein\_id="AAK16945.1"  
 /db\_xref="GI:13259574"  
 /translation="MVDVDNGVGFYSFSAUKDRHTAVEVNTDPQNGWQTSVELVED FEATLRYGRNLLKMDAFGCTSRGQAHRAGLWIKTGLLETQTVDFTLGSQGLRHTPGD

```
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

Alignment Scores:
Pred. No.: 34 Length: 297816
Score: 128.50 Matches: 49
Percent Similarity: 38.59% Conservative: 22
Best Local Similarity: 26.63% Mismatches: 66
Query Match: 16.54% Indels: 47
DB: 1 Gaps: 5

US-09-847-539a-6 (1-159) x AP002553 (1-297816)
Qy 15 GlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlu-----LysLeuAlaLeu 30
Db 83710 GGTACGCTGAATGATTTCTCGTGCCATGACGAGGATGATCCCTCGGAGGCACTG 83769
Qy 31 ArgAsnGluGluArgAlaLeuAspGluLeuLysGlnAlaLeuGluAspLysGlu--- 49
Db 83770 CGTGGTTTGAACAGATGTTGGAAGAAGACGCGCTCAGCAGAGGAGCGCAAGAAGAT 83829
Qy 49 ----- 49
Db 83830 GCGGAGAACGACAGACGTCGCGGAGAAATGCGGCAATATCAGCCAGTAAGCGGAAACG 83889
Qy 50 -----AlaThrAlaLeuAlaLeuAlaLeuSerSerAspAlaLeuGluAlaLeu 65
Db 83890 AGCCGCCAAATGCTGATCTTCAGCAGAGGATGTCGAGTCAGCCCGCAGCGCGCA 83949
Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAspAsnAla 85
Db 83950 GAAAGTCAGCCTCTGCAAGAGAGTCAGAGAGCGCTCTGCTCAGCCTCTGAGGCC 84009
Qy 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAla 105
Db 84010 GCTCAAAAGCCAGTGTAGTATTACAAAGTCACACAGATCCGAGTTGTTCANAAAGACG 84069
Qy 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125
Db 84070 GCAGAAAGTCAGCCGGTATGTCAGCCAGGATGCA----- 84105
Qy 126 AlaLeuAspValLysThrLysAspThrLysProValValLysLysGluGluArgGln 145
Db 84106 AGCACTCAACAGAAAAGCCGGAATCAGAGAAAGCCACAGTCAGCGGAACAAGC 84165
Qy 146 Asn-----ValAsnThrLeuProThr-----GlyGlu 155
Db 84166 AGAATACGCGGAGAACGCGGTAAACAGAAATTCCACCGTGTGTGGCGCTCCCGACCA 84225
Qy 156 GluSerAsnPro 159
Db 84226 AAGGGGAACCG 84237

RESULT 20
PMU19361
LOCUS PMU19361 3763 bp mRNA linear VRT 28-OCT-1995
DEFINITION Petromyzon marinus neurofilament subunit NF-180 mRNA, complete cds.
ACCESSION U19361
VERSION U19361.1 GI:632548
KEYWORDS sea lamprey.
SOURCE Petromyzon marinus
ORGANISM Petromyzon marinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
Petromyzontiformes; Petromyzontidae; Petromyzon.
1 (bases 1 to 3763)
JACOBS,A.J., Kamholz,J. and Seizer,M.E.
The single lamprey neurofilament subunit (NF-180) lacks
multiphosphorylation repeats and is expressed selectively in
projection neurons

JOURNAL Brain Res. Mol. Brain Res. 29 (1), 43-52 (1995)
MEDLINE 95287814
REFERENCE 2 (bases 1 to 3763)
AUTHORS Jacobs,A.J.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-1995) Alan J. Jacobs, Department of Neuroscience,
University of Pennsylvania, 452 Stemmler Hall, 3600 Hamilton Walk,
Philadelphia, PA 19104-6087, USA
FEATURES
Location/Qualifiers
1..3763
/organism="Petromyzon marinus"
/db_xref="taxon:7757"
/clone="clones LIF5, LIF13, and LIF22"
/tissue_type="central nervous system"
/clone_lib="Lambda gt11 cDNA library from larval lamprey
CNS"
/dev_stage="larva"
122..3454
/note="180 kDa neurofilament subunit"
/codon_start=1
/evidence=experimental
/product="NF-180"
/protein_id="AAA80106.1"
/db_xref="GI:632549"
/translacion="MSYTESVTCSYRAYADASRGSKRLLSLTPSSGFRSOSWCRGS
SGGSGSLVSLQRRPNYASLSHGYSYTPALNSAADSLDPHSQSVFSGDFKISRNEKQ
LVDDINDRFAGYIEKHVDLEQKNKELEISAYRKOSGPARGGVSDVYEQEIKELR
DQDIDDEKTKTQVQIEHLDDEIQRLREKTDDELRNTEALINAFKKNVDITSLV
RMDKRTQSLDEITFLKKNHEEVDLELLAQTSSTVSVERKDFAVPEITAAALREIR
GOLGOSARNIETAEEMFKGFSQLEAEOONDAIRSAKEETEHRRKLQMRCTELD
ALAGTKESLEROLSEMERHOSQDGNLQDAQOLENLNRTKEMARHLREYODLLNV
KHALDIEIAIRKLLGGEIRYSSGFLPTTPKPPKAPSAKPKAAKVEKKVKKPEIK
VESEPISQDLDLELAQVEEVAEAAKAPVSAEKDEEEEEEEEEEEEEEEEEEEEE
EDGRKEGEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
GEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
AEVEGEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
AEVEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
EDVEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
AEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
EPAKAPKPAKAPKVPKSEPEPEKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK
EPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE
EKEDDEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE
DKAEPKAKHSPVERKPIKEIAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK
EIQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ
IEITVSTTKKHKRAPVDSK"
BASE COUNT 1177 a 918 c 1248 g 420 t
ORIGIN
Alignment Scores:
Pred. No.: 1..1 Length: 3763
Score: 122.50 Matches: 37
Percent Similarity: 51.59% Conservative: 28
Best Local Similarity: 29.37% Mismatches: 50
Query Match: 15.77% Indels: 11
DB: 5 Gaps: 5

US-09-847-539a-6 (1-159) x PMU19361 (1-3763)
Qy 38 AspGluLeuLysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 57
Db 2234 GACCAACAGAGAGAGAGAGGTTAAAGAGAGAGAGGTCAC----- 2275
Qy 58 SerSerAspAlaLeuGluAlaLeuAlaAsp---GlnThrAspAlaLeuGlnSerGlu 76
Db 2276 AAGTCAGATGCTGAAGAGAGCTGAGGCTGAAGAGAGAGAGGCTGCAAGTCAGAG 2335
Qy 77 AlaAlaValValLysAlaAspAsnAlaAla-----SerAspAlaLeuGluAlaLeu 94
Db 2336 GAAGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2395
Qy 95 AspGlnThrAspAlaLeuGlnSerGlu---GluAlaGluValValGlnSerAspAsnAla 113
Db 2396 GAAGAAACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2455
```

MEDLINE REFERENCE AUTHORS	21156231 5 (bases 1 to 297816) Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.	gene CDS	NGPRFDLLSMLIAVRNLGVALLPRAIQLDLDSDGMWIPCDVPIRTGNRFIMTWOE EKSDSPHLQOFREWLAKSVYPOEM" /gene="ECs0756" complement(3720..4322) /gene="ECs0756" complement(3720..4322) /note="probable cob(I)alamin adenosyltransferase, similar to cob(I)alamin adenosyltransferases (corrinoid adenosyltransferases) e.g. [Escherichia coli] gi 115148 sp P13040 BTUR_ECOLI percent identity 67 in 200 aa"
	TITLE JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047) genome project.		
FEATURES SOURCE	Location/Qualifiers 1..297816 /organism="Escherichia coli O157:H7" /strain="O157:H7" /sub_strain="RIMD 0509952" /db_xref="taxon:83334" 79..1245 /gene="ECs0753" 79..1245 /note="similar to SUCC_ECOLI gi 1786948 percent identity 100 in 388 aa (Conserved in E.coli K-12)." /codon_start=1 /transl_table=11 /evidence=not_experimental /product="succinyl-CoA synthetase beta subunit" /protein_id="BAB34176.1" /db_xref="GI:13360212" /translation="MNLHEYQAKOLFARYGLPAPGVACTTPREAEAAKIGAGPMV VKQVHAGRGKAGGVKVVNSKEDIRAFENMLGKRLVYTOTDANGQFVNQILVEAAT DIAKELYGAGVDRSSRRVFMSTEGGVEIKVAEETPHLIHKVALDPLTGPMPYQG RELAKFLGLEKLVQOFTKIFMGLIATIFLERDLALIEINPLVITKQGDLCIDGLKLG DGNALFPQDLRENRQSOEDPREAQAAQWELNVALDNGICVAGLMTGMDIV KLHGEPANFLDVGSGATKERVTEAFKLIISDDKVAVLNIEGIVRCDLIDAGIIG AAVSGVNVVNVVRELGNAELGAKKLADSLNLTIAAKGLTDAAGQVVAVEG" 1245..2114 /gene="ECs0754" 1245..2114 /note="ECs0754" /note="similar to SUCD_ECOLI gi 1786949 percent identity 100 in 289 aa (Conserved in E.coli K-12)." /codon_start=1 /transl_table=11 /evidence=not_experimental /product="succinyl-CoA synthetase alpha subunit" /protein_id="BAB34177.1" /db_xref="GI:13360213" /translation="MSLIDKNTKVICQFTGSGTFSQAIAYGTMVGGVTPGKG GTHLGLPFTVREAVYATGATASVIYVAPFCKDSILEAIDAGIKLITITIGIPT LDMLTVKLDKDEAGVRMIGPCGVITPGECKIGIQPGHIHKPGKGVIVRSGLTYE AVKQTDYFGQSTCVGIGDPIPGSNFIDILEMFKDPQTEATVMEIGEGSAEEA AAVYKEHVTPVGVYIAGVTA PKGRMGHAGAI IAGCKGTADERFAALEAAGVKTVS LADIGEALKTVLK" 2781..3686 /gene="ECs0755" 2781..3686 /note="ECs0755" /note="probable transcriptional regulator, similar to transcriptional regulators e.g. glycine cleavage system transcription activator (gcv operon activator) - Escherichia coli gi 417043 sp P32064 GCVA_ECOLI percent identity 31 in 300 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative transcription regulator" /protein_id="BAB34178.1" /db_xref="GI:13360214" /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALQSAVROVSAL REFLTPFHSKRIIFLNDTKYLVKIVETLAKLERDNTIMTWQPTVOVIELAVN PTFETHMLIKHETKLHPDLIVNIHSLANNQDPLAREYDAVIMRENFCAPWAEVY LFEELIPVCSGSLNMSDQKLSVAELLTELPHLHQSTRITGNEENFALSCVSPLVN		
gene CDS	/codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative cob(I)alamin adenosyltransferase" /protein_id="BAB34179.1" /db_xref="GI:13360215" /translation="MEARISTEHRVQOQKLEQVDTRVAATEKKGILIVFCNGKG KSTAATGTVTRVAGHGTGVQAQIKQWNGEYNLLOPQVEFHNGTQFTWETQNR QADIDAAKVSNGSKRLADKRYDLVLDLYDELTYMLAYHLDTEEVIASLRPQAQSV IVTGRGCHSQILKMDTVSEIRPVVKHAFDNGIOAQPQIDW" complement(4332..5984) /gene="ECs0757" complement(4332..5984) /note="probable fumarate hydratase, similar to fumarate hydratases e.g. fumarate hydratase class I, aerobic (fumarase) - Escherichia coli gi 1205981 sp P00923 FUMA_ECOLI percent identity 68 in 545 aa"	gene CDS	/codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative fumarate hydratase" /protein_id="BAB34180.1" /db_xref="GI:13360216" /translation="MSKPFVQEPFLQNKQGTETVLSQDHTVTELDGDEVKIAIE ALTLLARQFYEAFFFLRSALHQVAILNDQOASSNDKYVALQLLNAEVSAGVLP NCQDTGATIVASGQQIWTGGNDAAEALSKGIVTFQENNLRFSONAPLDMYTEVNP TNLPAQIDISAVAGDEVHFLCVNKGSGSANKAALYQETKSLQPEKLTAFLEKMKSL GTACPPYHIAFVVGSLASDQTLKIAKLASTKYVDNLPTSGNEGOAFRDIETLEKVL EASQOFGIQAQFGSKYFAHDIRVIRLPHSGSCPIANALSCSADRNIAKINKHGIL EKLEHPCQIPASLREENHAQVQLDNLNRLPRDYMQDLARLPVGTIVSVLSGPIVAR DIAHAKIKARLDGSEPMPEYLKHIIHYIYAGPAKTPENNAACSGTPTGGRMDDYIDTF QAAGSLVMSLKGNSRQVTDACHKHGFLNLSIGGAAALLAQEYVKSRLCLEYPELG MEAVNMVENLPAFILLVDDKGNFFSQFEQQRHRCASCAPAGH" complement(6092..7372) /gene="ECs0758" complement(6092..7372) /note="probable transport protein, similar to glutamate/aspartate transport proteins (proton glutamate symport proteins) e.g. [Bacillus stearothermophilus] gi 1214671 sp P24943 GLTT_BACST percent identity 38 in 416 aa, also similar to C4-dicarboxylate transpor"
gene CDS	/codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative transcription regulator" /protein_id="BAB34178.1" /db_xref="GI:13360217" /translation="MKKISLTTMTLLALVLMGIMGVVNNNTASPETAKLYAQEISIFT TIFURLKMITAPLVVSTLVVGTAKMGDAKALGRIFSKTLFLFICASULLSIALGLITV NFMPTGTGINFVAGETGVAAEPTLKVFIISHAPTSIVDAAMHNEILQIVVFSI FLAGSLTAIGEKSAIIVHAEDLSLAHMLKLTGYVMLPALTIVFAAISALIAERGLAVM VSAGIFMGEYFTNLLVLLIGLAIYVVGPIRRLTRALSEPALLAFTTSSSAAFP GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMACTSPATFIAQCNIHUSIGSQITM LLILMLTSGMAGVPRASVMVIAATNQFINPEAGLILLMGVDFLDMGRSATVMN AMGAAMYSRWEGEHFEGECRCGKALKPNESNALP" complement(7533..7850) /gene="ECs0759" complement(7533..7850) /note="ECs0759"	gene CDS	/codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative transport protein" /protein_id="BAB34181.1" /db_xref="GI:13360217" /translation="MKKISLTTMTLLALVLMGIMGVVNNNTASPETAKLYAQEISIFT TIFURLKMITAPLVVSTLVVGTAKMGDAKALGRIFSKTLFLFICASULLSIALGLITV NFMPTGTGINFVAGETGVAAEPTLKVFIISHAPTSIVDAAMHNEILQIVVFSI FLAGSLTAIGEKSAIIVHAEDLSLAHMLKLTGYVMLPALTIVFAAISALIAERGLAVM VSAGIFMGEYFTNLLVLLIGLAIYVVGPIRRLTRALSEPALLAFTTSSSAAFP GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMACTSPATFIAQCNIHUSIGSQITM LLILMLTSGMAGVPRASVMVIAATNQFINPEAGLILLMGVDFLDMGRSATVMN AMGAAMYSRWEGEHFEGECRCGKALKPNESNALP" complement(7533..7850) /gene="ECs0759" complement(7533..7850) /note="ECs0759"

```

/db_xref="GI:12513757"
/translation="WLSPYŠVNLGCSWNSLTLNRLSPDNRLVLSVRDAAYHSDNGAQQ
KGNRYRYATNDKFCVTPRESHSGCFTNLLHRLGWPRGEISRKIEVLMNLASPVSAAM
KGVHNSRPDPDPPDYADPPLPSVDYINRLSPGVNIGKGNVAYVEDADATKVLKIM
FTTOSNEEVTSVRCFNQYAGSAEKIYGNNGDIIGIRMDKINGESLNLISLPAQ
ASHAIYDMFDRLEQKILPVDTTETNVLVYDRAKNFNPDISSVNVSDRSWSESQIMQ
SVHGGKODLISVLSKI"
gene
6462..7160
/genes="20990"
6462..7160
/genes="20990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="WRPTSLNVLHQSSRSSMSDIDIESLVKASSVQWIKNPQLRP
QQTIDINIQQIEAALDKIGSTETGRVLLNAIESIRLKSSETVVIHNSRLGVMAHRD
IDAEHNRGTGSDPHCNLNAVEYPCGEGISVDFHATIVFHELLHVFHNLGRSLKVES
SRAESQKSPLELEARTVGLGAFSEEVLSENKFHEIGMPRTSYPPXDSALIHONT
VSLGQQVRLHPLL"
gene
complement(7668..8144)
/genes="ybhb"
/notes="20992"
complement(7668..8144)
/genes="ybhb"
/function="orf; Unknown function"
/notes="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MCL655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="WKLISNDRLDGDKLPHRVFNGMGYDGNISPHLAWDDVPAGTK
SFVYCYDPDATGGGWHVWVNLPAOTRVLVLPQGGSLVAMPDGVLTQTRDFDKTG
YDGAAPPKGETHRYIFTVHALDVRIDVDEGASGAVGFNVHFHSLASASITAMFS"
gene
complement(8203..9492)
/genes="bloa"
/notes="20993"
complement(8203..9492)

Alignment Scores:
Pred. No.: 1.41 Length: 13501
Score: 128.50 Matches: 49
Percent Similarity: 38.59% Conservative: 22
Best Local Similarity: 26.63% Mismatches: 56
Query Match: 16.54% Indels: 47
DB: 1 Gaps: 5

US-09-847-539A-6 (1-159) x AE005258 (1-13501)
QY 15 GlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlu-----LysLeuAlaLeu 30
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 977 GGTACGCTGAATGATTTCTCGGTGCATGACGAGGAGTGATGCCGTCGGAGGCACTG 1036

QY 31 ArgAsnGluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGlu--- 49
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1037 CGTCGTTTGAACAGATGTTGGNAGAGCAGCCGTCACGACGAGGCGGAAGAAGAT 1096

QY 49 ----- 49

Db 1097 GCCGAGAACGACGACGTCGCCGAGGAATGCCGGCATATCAGCCAGTAAGCGGAAGCG 1156

QY 50 -----AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu 65
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1157 AGCGCGCAAAATGCTGATCTACTCAGCAGAGGATGTCATCGGATCGACGGCGGCGCA 1216

QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 85

```

```

Db 1217 GAAAGTGCAGCCCTCTGCAAGAAGTGCAGGAAGCGCTCTCGTCTCAGCGCC 1276
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1277 GCTCAAAAAGCCAGTGAAGTGCATTAACAAAGTGCACAGATGCCAGTGTCTCAAAAAGACG 1336
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125
::: ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1337 GCAGAAAGTGCAGCCGTAATGCAGCCAGGATGCA----- 1372
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 AlaLeuAspValLysThrLysAspThrLysProValLysLysGluGluArgGln 145
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1373 ACGACCTCAACAGAAAAGCCCGGAATCAGCAGAAAGCCACAGTCAGCGGAACAAAGC 1432
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 146 Asn-----ValAsnThrLeuProThrThr-----GlyGlu 155
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1433 AGAATAGCGCGGAAGAGCGCGTAACAGAAATTCACACCGTGTGGGGCTCCCGGACCA 1492
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 156 GluSerAsnPro 159
::: |||
Db 1493 AAGGGGGAACCG 1504

RESULT 19
LOCUS AP002553 297816 bp DNA linear BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C., and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
REFERENCE 2 (sites)
AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
REFERENCE 3 (sites)
AUTHORS Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
REFERENCE 4 (sites)
AUTHORS Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)

```

```
Escherichia.
1 (bases 1 to 13501)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A., and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
PUBMED
11206551
2 (bases 1 to 13501)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A., and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
Location/Qualifiers
1..13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
misc_feature
<1..7576
/note="O-island #36; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655; Cryptic prophage
CP-933k; includes one copy of the 13 bp direct repeat that
flanks the prophage"
66..665
/gene="lomk"
/note="20981"
66..665
/gene="lomk"
/function="putative membrane; Other or unknown (Phage or
Prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
residues 1 to 199 of 199 from Genpept 118 :
gi|7532789|gb|AA663231.1|AF151091_2 (AF151091) lom
[prophage P-Eiba]"
/codon_start=1
/transl_table=1
/product="putative outer membrane protein of prophage
CP-933k"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKVCAAILSAALCLAVSGVPAAWASEHQSTLSAGYLHASTDAPG
SDDLINGINVKRYEFTDPLGLITSPSVANAEDKQTHYSDTRWHEDEVNRFWSVMAG
PSVRVNEFSAXAMAGVAYSRVTSFGDYFRVTDNKRKTHDVLIGSDDARYSNTSLAW
GAGVQNPESVAVDVAVEXSGSDWRDTGFIYGVGYKF"
719..2041
/gene="20982"
719..2041
/gene="20982"
/function="putative structure; Structural component (Phage
or Prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical
to residues 381 to 645 of 645 from Genpept 118 :
gi|4585436|gb|AAD25464.1|AF125520.59 (AF125520) putative
tail fiber protein [Bacteriophage 933w]"
/codon_start=1
/transl_table=1
/product="putative tail component of prophage CP-933k"
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGNAAVOISGVLKDGACKPIONCTIQIKARNSTXYVYNTVASE
NPDENGRYSMDVEYQYSVTLVVGFPFPHAGTISVYEDSQPGLNDFLGAMTEDDAR
PEALRRFQMVEEAARHAEAKNAGEAETSARNAGISAKSAEASAAANADTSAEDASE
SARQAESAASAKKSEBASSSSAESAOKASESLOSATDAELSKKTKAESAGNAARDA
TTSTEKARSAESAQSAESQRIAAEDAVNRITPYVPPGPKGEPGAGPGKDGKE
RGDTGPAGATGEPGGDTGAGPQGPGRGDEGTGLTGAGPGQPGKGTGAAGPA
GPOGPKGTGAAGPVGATGPOGPKGTGCTGCTGAGPGLIETNSYGFPGTDCALI
TGLTFLDPKDATQVQMFQHLQVRFEGDPMQDVKGLDEVGSDTGRTE"
2004..2312
/gene="20984"
2004..2312
/gene="20984"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from Genpept 118 :
gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MKWAVIOALENDMNILRLKMQSLCGCKHDDCENGSRSLTAQLR
LGPADILDESDENGIIEQDRVITQVWILDADKKQIQCVVRPLQILRADGTWENIGMK"
2489..3469
/gene="20985"
2489..3469
/gene="20985"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from Genpept 118 :
gi|6960367|gb|IAAF33527.1| (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPRTTFHNSVNIQSSPCQTVSFAGKEVELKVIDEKTPILF
QWEPNPERYKKDEVPVINTKQHPYLDNVTNAARIESDRMIGIFVDGFSVNOKTAFS
KLEDFENVMIIYREDVDVDSMYDRKSLDIYHDIICEQRLTEDRDEYLLNLBKELR
EISKAQSLISMAYAKRNHAWDFEFLNALLAGEIFRCTYTNKHNLSFGEGCIYLD
MDMLTKLGTIYAPDGMISHVDRNDSNIENSALIVNRSNHPALLEGSLFPMHSKVD
AHPTYDGLGKGVKYYFNFTPLHNYNHFCDFIEFNHPINIMNTSQYTCSSW"
3530..4522
/gene="20986"
3530..4522
/gene="20986"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="WKIPSLQSNFNSAPAGYSAPIAPNRAENAYADYVLDIGKRPL
SAADLSNVYSVIRWAKPSLQSLRSLIDQHTVDMTGNVLDAALSRQTFDRAVSYGTHNEK
VHIGCIKYREYELNESSVKIDDIQSITCNELVEYDVQGPPIFCIACEAGNDNEEY
VSFSVADTDYSMPQBGELHETIHHVTGSSDPGSDSNIELGPTTEILARRVAQELG
WSVPDFPGYAEPAERLAHLRLNALRQAAHRBEENRAFFERLGTISDRYEASPDFT
EYSVSNIGYGIQOHDHDPGLAINDNLQDANQIQLYHGAPYITFTFGDVDRHNNQ"
5350..6231
/gene="20989"
5350..6231
/gene="20989"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55142.1"
```

```
JOURNAL Infect. Immun. 60, 3601-3608 (1992)
MEDLINE 92363555
FEATURES
    source
        Location/Qualifiers
            1..969
            /organism="Streptococcus canis"
            /isolate="DG12"
            /db_xref="taxon:1329"
            <1..>969
            /codon_start=1
            /transl_table=11
            /product="albumin-binding protein"
            /protein_id="AAA26847.1"
            /db_xref="GI:15355"
            /translation="SVLVGANTVSAQVTTQAQAKARETALASIAETESSIKDKAQLD
            SYKERIKASVDREITELSAQADQIVSAQADNEITKABEDSSKAWEAADQANTAKA
            EADELAKESKESDAWEAKAALDQAKAALKEFDRIYGVSNYYKNLINKAKTVEGTIMEL
            QAQVESAKKARISATDGLSDFLKSQTSADETLKS IKLSEAKEMAIREDLAQGVDF
            YKNKINNAKTVEGVVAKDLILNSLAKPEVPEAKPEVKPEKPEVPEVKPEV
            KPEVKPEVKPEVKPEVKPEVKPENKPAEVKDEKLLPSTGESV"
    CDS
        359 a 168 c 219 g 223 t
        BASE COUNT 359 a 168 c 219 g 223 t
        ORIGIN
            Alignment Scores:
            Pred. No.: 0.016 Length: 969
            Score: 138.50 Matches: 39
            Percent Similarity: 55.65% Conservative: 25
            Best Local Similarity: 33.91% Mismatches: 36
            Query Match: 17.82% Indels: 15
            DB: 1 Gaps: 4
            US-09-847-539A-6 (1-159) x STRABP (1-969)
            QY 42 LysGluAlaIleGluAspLysGluAlaThrAlaIleGlu-----AlaAlaSer 58
            Db 100 GAATCATCTATCAAGACAAACACAGATTAGATTTCAGTAAAGAAAAATTAAGCTTCT 159
            QY 59 SerAsp-----AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 73
            Db 160 GTTGATAGACAGATACAGATTGTCGGCCGACAGCAGACAGATTGTTCTGCACAA 219
            QY 74 SerGluGluAlaAlaValValLysAlaAspAlaAlaSerAspAlaLeuGluAlaLeu 93
            Db 220 GCAGATAACCAAGCTATTACTAAGCAGAGAAGAGATTCTTCCAAGGCATGGGAAGCTGCT 279
            QY 94 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAla 113
            Db 280 CGGGATCAACCAATACTGCTAAAGCAGAGAAGCAGATCACTGCAAGCGCAAGAAAGAA 339
            QY 114 AlaSerAspAlaThrGluLysAlaAlaThrProileAlaLeuAspValLysLysThrLys 133
            Db 340 TCATCAGATGCTTGGCAGAGGCTGCT-----GCTTTAGATCAAGCTAAG----- 384
            QY 134 AspThrLysProValValLysLysGluGluArgGlnAlaValAsn 148
            Db 385 -----CAAGCTGCTCTAAAGAATTCGATCGTTATGTGTGAGC 423
            RESULT 17
            PFAS7
            LOCUS 3838 bp DNA linear INV 26-APR-1993
            DEFINITION Plasmodium falciparum (isolate NF7) S antigen gene, complete cds.
            ACCESSION M10130
            VERSION M10130.1 GI:160670
            KEYWORDS S antigen.
            SOURCE P.falciparum (isolate NF7) DNA, clone NF7.S.
            ORGANISM Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
            REFERENCE 1 (bases 1 to 3838)
            AUTHORS Cowan,A.F., Saint,R.B., Coppel,R.L., Brown,G.V., Anders,R.F. and Kemp,D.J.
            TITLE Conserved sequences flank variable tandem repeats in two S-antigen genes of Plasmodium falciparum
            JOURNAL Cell 40, 775-783 (1985)
            MEDLINE 85176931
            COMMENT Twenty degenerate tandem repeats of twentyfour base pairs each start at position 953 and end at 1432.
            FEATURES
                source
                    Location/Qualifiers
                        1..3838
                        /organism="Plasmodium falciparum"
                        /db_xref="taxon:5833"
                        665..733
                        /note="S antigen signal peptide"
                        665..1594
                        /note="S antigen precursor"
                        /codon_start=1
                        /protein_id="AAA29758.1"
                        /db_xref="GI:160671"
                        /translation="MNRILSVSEYLFELIYIKTYGKVKNTDQELSNIVGTNYLNRN
                        GFUNGKNGKNTYEDLOEGBEGENDEEHSNDEENETIIVGDEAPKSDAEAK
                        LKSDAEALKSDEAEARKSDEAEALKSDEAEARKSDEAEALKSDEAEARKSDEAEARK
                        SDAEARKSDEAEARKSDEAEARKSDEAEARKSDEAEARKSDEAEARKSDEAEARKS
                        EAPARKSDEAEARKSDEAEARKSDEAEARKSDEAEARKSDEAEARKSDEAEARKS
                        EGPKGTVGGPGSGGESHKSKKSKIMMLILM"
                sig_peptide
                CDS
                    mat_peptide
                        734..1591
                        /product="S antigen"
                        BASE COUNT 1601 a 270 c 504 g 1463 t
                        ORIGIN 1 bp upstream of EcoRI site.
                        Alignment Scores:
                        Pred. No.: 0.228 Length: 3838
                        Score: 131.50 Matches: 43
                        Percent Similarity: 57.14% Conservative: 29
                        Best Local Similarity: 34.13% Mismatches: 45
                        Query Match: 16.92% Indels: 9
                        DB: 3 Gaps: 6
                        US-09-847-539A-6 (1-159) x PFAS7 (1-3838)
                        QY 26 GluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIle 45
                        Db 959 GAGCAGAGCAGCACTAAAAGCTGATGAG-----GCAGAGGCACCTAAAAGCTGATGAGGCA 1012
                        QY 46 GluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAla--- 64
                        Db 1013 GAGCGACGAAAAAGTCTGATGAGCA--GAGCGACATAAAAGTCTGATGAGGAGGCACGA 1069
                        QY 65 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsn 84
                        Db 1070 AAAAGTGTGAGCAGGAGGCATAAAAGTGTGATGAGCAGGAGGCATAAAAGTGTGAG 1129
                        QY 85 Ala-----AlaSerAspAlaLeuGluAlaLeu---AlaAspGlnThrAspAlaLeu 100
                        Db 1130 GCAGAGCAGCAAAAAAGTCTGATGAGGAGGCAGGCACCTAAAAGTCTGATGAGGAGGCACCTA 1189
                        QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaThrGluLys 120
                        Db 1190 AAAAGTGTGAGCAGGAGGCACGAAAAAGTCTGATGAGCAGGAGGCACGAAAAAGTGTGAG 1249
                        QY 121 AlaAlaThrProileAlaLeuAspValLysLysThrLys--AspThrLysProValVal 139
                        Db 1250 GCAGAGCAGCAAAAAAGTGTGAGGAGGCAGGAGGCACGAAAAAGTGTGAGGAGGCACCTA 1309
                        QY 140 LysLysGluGluArgGln 145
                        Db 1310 AAAAGTGTGAGCAGGAG 1327
            RESULT 18
            AE005258
            LOCUS 13501 bp DNA linear BCT 21-MAR-2001
            DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82 of 155.
            ACCESSION AE005258
            VERSION AE005258.1 GI:12513751
            KEYWORDS
            SOURCE Escherichia coli O157:H7 EDL933.
            ORGANISM Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
```

US-09-847-539A-6 (1-159) x SSCMLB1 (1-1380)		Pred. No.:	0.012	Length:	798
Qy	19 AsnLeuLeuGlyAsnAla	Score:	139.00	Matches:	58
Db	601 AATCTTTTGGGAATCTAAAGATCAATTAGATAAACTTACTACAGAAAAGTGTG	Percent Similarity:	35.65%	Conservative:	24
Qy	31 ArgAsnGluGluArgAlaIleAspGluLeuLysLys	Best Local Similarity:	25.22%	Mismatches:	64
Db	661 TCTAAAGATAAAGANGCATAGTAGTAAAGAAATATCTGTAAGCAAGTCGTAACCGT	Query Match:	17.89%	Indels:	84
Qy	43	DB:	6	Gaps:	8
Qy	721 ACAACCGTGATTTAGAGCGAGCTCGTGACGCTAAGAACGCCACAGAGATAGCT	US-09-847-539A-6 (1-159) x I09107 (1-798)			
Qy	58 SerSerAspAla	Qy	4	ProIleGluGlnProArgIleIleProAsnGlyThrLeuThrAsnLeuLeuGlyAsn	23
Db	781 GAAACAAACCTAAAGCTTGACAAGCTTGAAGAGAAAACAAAGTCTTGTGAAGCCAGCGGT	Db	100	CCAATCGAAGATACCCCAATTTATCGTAAATGCTGTGAATTAACATACTTCTGGGAAT	159
Qy	67 AspGlnThr	Qy	24	AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGlu	39
Db	841 AAACGTCACAAACCGTGATTTAGAGCGAGCTCGTGACGCTAAGAAAGCCACAGAACGAGAA	Db	160	TCAGAGACAAACACTGGCTTTGCTAATGAAGAGAGTGTACAGCTGGGTACCCATTACCT	219
Qy	79 ValValLysAlaAsp	Qy	40	LeuLysLysGlnAlaIleGluAspLysGluAlaThr	51
Db	901 TTGGCTAAAGCTAATGAGTAAACCAATCTTAGAAGCAAGCCGACACGACCAACCGT	Db	220	AAAGCTGACACTTACAAATTAATCCCTTAATGCTAAACATCTGTAAGGCGAAACACTACT	279
Qy	84	Qy	52	ThrAlaIleGluAlaIleSerSerAsp	61
Db	961 GATTTAGAGCGAGCTCGTGACGCTAAGAAAGCAGTGTGTCAGAACTTAA	Db	280	GAAGCTGTTGATCTGCTACTGCAGAAAAGCTTCAACAAATACGCTAACGACAAACCGT	339
Qy	95 AspGlnThrAspAlaLeuGlnSerGluGluAlaGluValGlnSerAspAsnAla	Qy	62	LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln	77
Db	1021 GCTCAAGCTGAGCGCTTAAAGAGCAATTTGGCTAAACAGCTCAAGAAATTTGAAAACTT	Db	340	GTTGACGGTGAATGGACTTACGACGATGCGACTAAGACGCTTTTACAGTTACTGAAAAACCA	399
Qy	115 SerAspAlaTyrGluLysAla	Qy	78	AlaValValLysAlaAsp	84
Db	1081 AAAGATCAAAAGAAAAGACCCAGCAAGCAACTCAAACTCCAGAAAACCCAGAGTACCT	Db	400	GAAGTGTGATGCTGCTGCTGAATTAACACCGCGGTGACAACTTACAAACTTGTATTAAAT	459
Qy	122	Qy	85	AlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu	104
Db	1141 GGTAAACCAAGCATGCTTTGGACAGGATTAACACCTGCTACACCAATCGCTAAGACAGACA	Db	460	GCTAAACATTTGAAGCGCAACCACTACTAAAGCAGTACGACGAGAACTGCAGAAAAA	519
Qy	130 LysLysThrLysAspThrLysProVal	Qy	105	AlaGluValValGlnSerAspAsnAlaIleSerAspAlaTyr	118
Db	1201 AAGAAACT---GATGTTAAACCTGCGCTAAGCAACATCGTGCCTACTGACGTTAAG	Db	520	GCCTTCAACAAATACGCTAAGCAGACACGCTGTTGATGCTGTTGGACTTATGATGATCG	579
Qy	141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsnPro	Qy	118	ACTAAGACCTTTACGGTAACTGAATGGTTACAGAGGTTCTTCGAGGTGATGCACCAACT	639
Db	1258 AAAGATGAGAGAAA-----CTTCATCACTGAGGAACTGTGAACCCA	Qy	119	Glulys-----AlaAlaThrProIle	125
RESULT 15		Db	640	GAACGAAAAAACCCAGCAAGTATCCCTCTTGTTCGCTTAACCTCCTGCAACTCCAATT	699
I09107		Qy	126	AlaLeuAspValLysLysThrLysAspThr-----LysProValVal	139
LOCUS	109107	Db	700	GCTAAAGATGACGCTAAGAAAGACGATACTAAGAAAGAGAGTGTATAAAACCCAGAGCT	759
DEFINITION	Sequence 4 from Patent WO 8810306.	Qy	140	LysLysGluGluArgGlnAsnValAsnThr	149
VERSION	109107.1	Db	760	AAGAAAGATGACGCTAAGAAAGCTGAAACT	789
KEYWORDS	GI:588179	RESULT 16			
SOURCE	Unknown.	STRABP			
ORGANISM	Unclassified.	LOCUS			
REFERENCE	1 (bases 1 to 798)	DEFINITION			
AUTHORS	Fahnestock,S.R.	STRABP			
TITLE	CLONED PROTEIN G VARIANT GENES AND THE PROTEIN G VARIANTS EXPRESSED THEREFROM	LOCUS			
JOURNAL	Patent: WO 8810306-A 4 29-DEC-1988;	DEFINITION			
FEATURES	Location/Qualifiers	STRABP			
source	1..798	LOCUS			
BASE COUNT	277 a 161 c 177 g 183 t	DEFINITION			
ORIGIN	/organism="unknown"	STRABP			
Alignment Scores:		LOCUS			

Pred. No.: 0.00155 Length: 1576  
Score: 154.50 Matches: 80  
Percent Similarity: 26.45% Conservative: 16  
Best Local Similarity: 22.04% Mismatches: 62  
Query Match: 19.88% Indels: 206  
DB: 1 Gaps: 7

US-09-847-539a-6 (1-159) x SG148PG (1-1576)

```
QY 1 ValAspSerProIleGluGlnProAtrGleIleProAsnGlyThrLeuThrAsnLeu 20
||||| 1576
Db 391 GTTGACTCAACATCGAAGATACCCCAATTATTGTAATGGTGGTAATTAACCTAATCTT 450
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
||||| 1576
Db 451 CTGGGAATTCAGACACACACTGGCTTTCGTAATGAAGAGAGTGTACAGCTGATTG 510
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaSerSerAsp 60
||||| 1576
Db 511 ACAGCAGCAGCGGTAGCCGATCTGTGCAGCAGCGCGCATGAAATGCTGGCGCAGCA- 569
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
||||| 1576
Db 570 GCTTGGGAAGCAGCGGCA---GCAGCAGATGCTCTAGCAAAAGCCAAAGCAGATGCCCTT 626
QY 81 LysAlaAspAsn----- 84
Db 627 AAGAATTCACATAGATGAATTTTACGTGCTTACCTAAGACTGACACTTACAAATTA 686
QY 85 -----AlaAlaSerAspAlaLeuGluAlaLeuAla 94
Db 687 ATCCTTAATGGTAAACATTGAAAGCGAAACAACACTACTGAAGCTGTTGATGCTGCTACT 746
QY 95 AspGln----- 96
Db 747 GCAGAAAAAGCTTCAAAACAATACGCTAAGCAACACGGTGTGACGGTGAATGCACTTAC 806
QY 96 ----- 96
Db 807 GACGATCGGACTAAGACCTTTACAGTTACTTGAACAAACCAGAAAGTGCATGCTGTGAA 866
QY 96 ----- 96
Db 867 TTAACACCGCGGTGACAACCTTACAACCTGTTATTAAATGGTAAACATTGAAAGCGGAA 926
QY 97 -----ThrAspAlaLeuGlnSerGluGluAlaGluVal-----GlnSer 110
Db 927 ACAACTACTGAAGCTGTTGATGCTGCTACTGCAGAAAAAGTCTTCAAAACAATACGCTAAC 986
QY 111 AspAsnAlaAlaSerAspAlaIleP----- 118
Db 987 GACAAACGGTGTTCACGGTGAATGGACTTACGACGATGCGGACTAAGACCTTTACAGTTACT 1046
QY 118 ----- 118
Db 1047 GAAAAACAGAGTATGATGCGTCTGTAATTAACACACCGCTGACAACTTACAAACTT 1106
QY 118 ----- 118
Db 1107 GTTATTAAATGGTAAACATTGAAAGCGGAAACAACACTACTAAGCAGTAGACGCGAGAACT 1166
QY 119 ---GluLysAla----- 121
Db 1167 GCAGAAAAACCCCTTCAAAACAATACGGTAAACGCAACCGTCTTGTGATGGTGTGGACTTAT 1226
QY 121 ----- 121
Db 1227 GATGATCGGACTAAGACCTTTACGTTAACTGAATAAGTTTACAGAGTTCTCTGCTGATGC 1286
QY 122 -----AlaIle 123
Db 1287 ACCAACTGAACCAAGAAAAACCAAGAACGATATCCCTCTTTGTTCCGTTAACTCCTGCAAC 1346
```

```
QY 123 rProIleAlaLeuAspValLysLysThrLysAspThr-----LysPr 137
||||| 1370 bp DNA linear BCT 12-JUN-1996
Db 1347 TCCAAATGCTAAAGATGACGCTAAGAAAGACGATACCTAAGAAAGACGATCTAAATAAAC 1406
QY 137 ovalValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSe 157
||||| 1370 bp DNA linear BCT 12-JUN-1996
Db 1407 AGAAGCTAAGAAAGACGCTAAGAAAGCTGAAACTCTTCTTACAACTGCTGAAGGAA 1466
QY 157 rAsnPro 159
Db 1467 CAACCCA 1473
RESULT 14
SSCMLB1
LOCUS
DEFINITION
Strepptococcus sp.serotype C mbl1 gene for multiple ligand-binding
protein 1.
ACCESSION
X84989
VERSION
GI:1403574
KEYWORDS
mbl1 gene; multiple ligand-binding protein.
SOURCE
Strepptococcus sp.
ORGANISM
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Strepptococcus.
REFERENCE
1 (bases 1 to 1380)
AUTHORS
Talav,S.R.
TITLE
Direct Submission
JOURNAL
Submitted (23-FEB-1995) S.R. Talav, GBF Gesellschaft
f. Biotechn.Forschung mbH, National Research Centre for Biotechn.,
Mascheroder Weg 1, D-38124 Braunschweig, FRG
REFERENCE
2 (bases 1 to 1380)
AUTHORS
Talav,S.R., Grammel,M.P. and Chhatwal,G.S.
TITLE
Structure of a group C streptococcal protein that binds to
fibrinogen, albumin and immunoglobulin G via overlapping modules
JOURNAL
Biochem. J. 315 (Pt 2), 577-582 (1996)
MEDLINE
96202013
FEATURES
Location/Qualifiers
source
1..1380
/organism="Streptococcus sp."
/strain="serotype C"
/isolate="C20"
/db_xref="taxon:1306"
61..168
/gene="mbl1"
61..1380
/gene="mbl1"
61..1380
/gene="mbl1"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="multiple ligand-binding protein 1"
/protein_id="CAA59349.1"
/db_xref="GI:1403574"
/db_xref="SPTREMBL:Q55105"
/translation="MARKKTKQYSIRKLSVGNASVLVATGLIGGATVSNSEYSDRIT
TGTSNGHITSSGIGSTIVGGIAKPSNIIGTGTAGTIGSGTGGTIVSGTSGSN
TASGTVAGACGTTHDGNTIASGVENKDLKEVAELTEANALSTLIDKTTIEKIK
SEYKKEIGCKELKLDHI NNLLGNKQDQDKLTTEKESLKDKEALDERNKVLEASR
TRNRLDEAARQAKATEAETNAKVDLEEKOVLEASRKRTNRDLNARADAKKA
TEAELAKANELNQILEASRTNRDLNARADAKKAVDAELAKLKAELKDELAKQA
QELKELKESKAPKAPKATQTPKPEVPGKPSMPKWTGLTPATPIAKDKKTDVPRPAKAN
MPTDVKKDEKPLPSTGTNTNPFPTAANGMAGMATAGVAVAGRKREN"
```

BASE COUNT 530 a 244 c 314 g 292 t

Alignment Scores:

Pred. No.:	0.0023	Length:	1380
Score:	151.50	Matches:	65
Percent Similarity:	37.66%	Conservative:	25
Best Local Similarity:	27.20%	Mismatches:	46
Query Match:	19.50%	Indels:	103
DB:	1	Gaps:	10



```
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
|||||
Db 505 CTGGGAATTCAGACACACACTGCTTGGCTAATGAAGAGAGTGCTACAGCTGATTG 564
|||||
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
|||||
Db 565 ACAGCAGCAGCGGTAGCCGATCTGTGCAGCAGCGCGAGCTGAANAATGCTGGGCAGCA 624
|||||
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVal 80
|||||
Db 625 GCTTGGGAAGCAGCGCGCA---GCAGCAGATGCTCTAGCAAAAGCCAAAGCAGATGCCCTT 681
|||||
Qy 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
|||||
Db 682 AAAGAATTCACAATAATGGAGTACGATCTATACAGAATCTAATCAACAATGCCAAA 741
|||||
Qy 96 GlnThrAspAlaLeuGlnSerGluGluAlaValValGlnSerAspAlaAla--- 114
|||||
Db 742 ACTGTTGAAGCGGTAAAGACCTTCAAGCACAAGTTGTTGAATCAGCGAAGAAAGCGGT 801
|||||
Qy 115 ---SerAspAlaTrpGlu-----LysAlaAlaThrProIleAlaLeu 127
|||||
Db 802 ATTTCAAGAGCAACAGATGGCTTATCTGATTTCTTGAATCACAACACCTGCTGAAGAT 861
|||||
Qy 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
|||||
Db 862 ACTGTTAAATCAATTAAGTATTAGCTGAAGCTTAAGCTTAGCTAACAGAGAA 912
|||||
RESULT 12
108536
LOCUS 108536 1950 bp linear PAT 02-DEC-1994
DEFINITION Sequence 2 from Patent WO 8705025.
ACCESSION 108536
VERSION 108536.1 GI:588754
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
BASE COUNT 706 a 324 c 397 g 523 t
ORIGIN
Alignment Scores:
Pred. No.: 7.29e-06 Length: 1950
Score: 186.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 23.94% Indels: 16
DB: 6 Gaps: 6
US-09-847-539A-6 (1-159) x 108536 (1-1950)
Qy 1 ValAspSerProIleGluGlnProhrgllelleProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
Db 677 GTTGATTCACCAATCGAAGATACCCCAATTTATTCGTAATGGTGTGCAATTAATCTT 736
|||||
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
|||||
Db 737 CTGGGAATTCAGACACACACTGCTTGGCTAATGAAGAGAGTGCTACAGCTGATTG 796
|||||
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
|||||
Db 797 ACAGCAGCAGCGGTAGCCGATCTGTGCAGCAGCGCGAGCTGAANAATGCTGGGCAGCA 856
|||||
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVal 80
|||||
```

```
Db 857 GCTTGGGAAGCAGCGCA---GCAGCAGATGCTCTAGCAAAAGCCAAAGCAGATGCCCTT 913
|||||
Qy 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
|||||
Db 914 AAAGAATTCACAATAATGGAGTACGATCTATACAGAATCTAATCAACAATGCCAAA 973
|||||
Qy 96 GlnThrAspAlaLeuGlnSerGluGluAlaValValGlnSerAspAlaAla--- 114
|||||
Db 974 ACTGTTGAAGCGGTAAAGACCTTCAAGCACAAGTTGTTGAATCAGCGAAGAAAGCGGT 1033
|||||
Qy 115 ---SerAspAlaTrpGlu-----LysAlaAlaThrProIleAlaLeu 127
|||||
Db 1034 ATTTCAAGAGCAACAGATGGCTTATCTGATTTCTTGAATCGCAAAACCTGCTGAAGAT 1093
|||||
Qy 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
|||||
Db 1094 ACTGTTAAATCAATTAAGTATTAGCTGAAGCTTAAGCTTAGCTAACAGAGAA 1144
|||||
RESULT 13
SG148PG
LOCUS 1576 bp DNA linear BCT 23-MAR-1993
DEFINITION Streptococcus G148 protein G' structural gene.
ACCESSION X53324
VERSION X53324.1 GI:288358
KEYWORDS Streptococcal Protein G.
SOURCE Streptococcus sp. 'group G'.
ORGANISM Streptococcus sp. 'group G'.
REFERENCE
AUTHORS Goward,C.R., Murphy,J.P., Atkinson,T. and Barstow,D.A.
TITLE Expression and purification of a truncated recombinant streptococcal protein G
JOURNAL Blochem. J. 267 (1), 171-177 (1990)
MEDLINE 90226312
FEATURES
source
1..1576
/organism="Streptococcus sp. 'group G'"
/strain="G148"
/db_xref="taxon:1320"
/clone="pspg3"
292..1262
/gene="Protein G'gene"
292..603
/gene="Protein G'gene"
note="Truncated gene"
/codon_start=1
/transl_table=11
/product="Protein G'"
/protein_id="CAA37409.1"
/db_xref="GI:288359"
/db_xref="SPTREMBL:Q54180"
/translation="MERKKVKYELRKSAGFLGSVAFLVGSTVFAVDSDIEDTPIL
RNGELTNLLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGQGLKQRQQOM
L"
705..1262
/gene="Protein G'gene"
note="Truncated gene; Start codon TTG not ATG"
/codon_start=1
/transl_table=11
/product="Protein G'"
/protein_id="CAA37410.1"
/db_xref="GI:581662"
/db_xref="SPTREMBL:Q54181"
/translation="MKGETTTEAVDAATAEKVKFOYANDNCVDGEWYDDATKTKFTVT
EKPEVIDASELTSPAVTYKLVIKNGTKLGETTTEAVDAATAEKVKFOYANDNGVDGEW
TYDATKTFVTEKPEVIDASELTSPAVTYKLVIKNGTKLGETTTRKVAETAETAEKFK
QYANDNGVDGWVYDDATKTKFTVE"
BASE COUNT 572 a 273 c 327 g 404 t
ORIGIN
Alignment Scores:
```



```

Db 797 ACAGCAGCAGCGGTAGCCGACTACTGTGGCAGCAGCGGCGAGCTGAAATGCTGGGCGAGCA 856
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVal 80
   ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 857 GCTTGGGAAGCAGCGGCA---GCAGCAGATGCTCTAGCAAAAGCCAAAGCAGATGCCCTT 913
Qy 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
   ||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 914 AAAGAAATTCACAAATATATGAGTACTATTACAGAAATCTATCAACAATGCCAAA 973
Qy 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAlaAla--- 114
   ||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 974 ACTGTTGAAGGATAAAGACCTTCAAGCACAGTGTGTAATCAGCGAAGAAAGCGCT 1033
Qy 115 ---SerAspAlaTrpGlu-----LysAlaAlaThrPrfilleAlaLeu 127
   ||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1034 ATTTCAAGAACACAGCTTATCTGATTTCTTGAATTCGCAACACCTGCTGAAGAT 1093
Qy 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
   ||||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1094 ACTGTTAAATCAATTCGAATAGCTGAAGCTAAAGTCTTAGCTAACACAGAA 1144

RESULT 9
SGSPG
LOCUS
DEFINITION Streptococcus sp. spg gene for protein G. linear BCT 09-OCT-1997
ACCESSION Y00428
VERSION Y00428.1 GI:47100
KEYWORDS G protein; IgG binding protein; spg gene.
SOURCE Streptococcus sp. GX7805.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 2384)
AUTHORS Fahnestock, S.R.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-1987) Fahnestock S.R., Genex Corp., 16020
Industrial Dr., Gaithersburg, MD 20877, USA
REFERENCE 2 (bases 1 to 2384)
AUTHORS Filpula, D., Alexander, P. and Fahnestock, S.R.
TITLE Nucleotide sequence of the protein G gene from Streptococcus
GX7805, and comparison to previously reported sequences
JOURNAL Nucleic Acids Res. 15 (17), 7210 (1987)
MEDLINE 88015586
FEATURES
    Location/Qualifiers
        1..2384
        /organism="Streptococcus sp. GX7805"
        /strain="GX7805"
        /db_xref="taxon:1325"
        /db_xref="taxon:1325"
        577..2358
        /gene="spg"
        /gene="spg"
        /gene="spg"
        /gene="spg"
        /codon_start=1
        /transl_table=11
        /product="protein G precursor"
        /protein_id="CAA68489.1"
        /db_xref="GI:47101"
        /db_xref="SWISS-PROT:P19909"
        /translation="MEKEKKVKYIFURKSAFLGSLVSAAPLVGSTVFVAVDSPTEDPTII
        RRGELTNLGNSETTLALRNEESATADLTAAAVADTVAAANAGAAWAAAD
        ALAKADALKEFNKYGVSDYKLNINNAKTVEGKDLQAOVSEAKARISEATDGL
        SDFLSQTPAEDTVKSIELAEKVLNRELDKYGVSDYKLNINNAKTVEGKDLQAO
        VSEAKARISEATDGLSDFLSQTPAEDTVKSIELAEKVLNRELDKYGVSDYKLN
        LINNAKTVEGKDLQAOVSEAKARISEATDGLSDFLSQTPAEDTVKSIELAEKVLN
        ANDNGVDEMTYDDATDEILALPRTDPTKLLNGKTLKGETTTERDAVDAKTEKPY
        DAATKEKTVKQYANDNGVDEMTYDDATKTFVTETKPEVIDASELTPATVTKLVING
        KTLKGETTTKAVDAEAKFKQYANDNGVDEMTYDDATKTFVTETKPEVIDASELTPATV
        EPEKPEASTPLPTPATPIAKDDAKDDTKEDAKKPEAKKEDAKKAEETLPTTGEES
        NPFFTAAALAVMAGALAVASKRED"
        676..2355
    mat_peptide

```

```

/gene="spg"
/product="protein G"
886..996
/note="A1 repeated sequence"
986..1435
/note="225 bp duplication"
/rpt_unit=1211..1435
1111..1221
/note="A2 repeated sequence"
1354..1446
/note="A3 repeated sequence"
1483..1647
/note="B1 repeated sequence"
1499..1918
/note="210 bp duplication"
/rpt_unit=1708..1918
1693..1857
/note="B2 repeated sequence"
1903..2067
/note="B3 repeated sequence"
BASE COUNT 868 a 408 c 482 g 626 t
ORIGIN
Alignment Scores:
Pred. No.: 7,51e-06 Length: 2384
Score: 187.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 24.07% Indels: 16
DB: 1 Gaps: 6
US-09-847-539A-6 (1-159) x SGSPG (1-2384)
Qy 1 ValAspSerProFlegGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 676 GTTGACTCACCACCAATGAGTACCCCAATATTCGTAATGGTGGTGAATTAACATCTT 735
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 736 CTGGGAATTCAGAGACACACTGGCTTTCGCTAATGAAGAGAGTGCTCAGCTGATTG 795
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerAsp 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 796 ACAGCAGCAGCGGTAGCCGATATCTGTGGCAGCAGCGGCGAGCTGAAATGCTGGGCGAGCA 855
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
   ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 856 GCTTGGGAAGCAGCGGCA---GCAGCAGATGCTCTAGCAAAAGCCAAAGCAGATGCCCTT 912
Qy 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
   ||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 913 AAAGAAATTCACAAATATGAGTACTATTACAGAAATCTATCAACAATGCCAAA 972
Qy 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAlaAla--- 114
   ||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 973 ACTGTTGAAGGCGTAAAGACCTTCAAGCACAGTGTGTAATCAGCGAAGAAAGCGCT 1032
Qy 115 ---SerAspAlaTrpGlu-----LysAlaAlaThrPrfilleAlaLeu 127
   ||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1033 ATTTCAAGAACACAGATGCTTATCTGATTTCTTGAATTCACAAACACCTGCTGAAGAT 1092
Qy 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
   ||||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1093 ACTGTTAAATCAATTCGAATAGCTGAAGCTAAAGTCTTAGCTAACACAGAA 1143
RESULT 10
LOCUS
DEFINITION Sequence 4 from Patent WO 8705025. linear PAT 02-DEC-1994
ACCESSION I08537
VERSION I08537.1 GI:588755
KEYWORDS Unknown.
SOURCE

```

```

|||||
481 GCTCGGTGTTAAAGCGGATACGCTGCTAGTGACACTTTAGAGCATTTGGCGGATCAA 540
Db

97 ThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAsp 116
Qy
|||||
541 ACAGACGCTTTACAAATCAGAGAAGAGCTGAAGTAGTTTCAATCAGATAACGCTGCTAGTGAC 600
Db

117 AlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysThrLysAspThrLys 136
Qy
|||||
601 GCCTGGGGAAGAGCAGCACTCCCAATCGCTTTAGATGTTAAGAAAATAAGATACAAAA 660
Db

137 ProValValLysLysGluGluAlaGlnAsnValAsnThrLeuProThrGlyGlu 155
Qy
|||||
661 CCTGTACTTAAAGAAAGAAAGACAAACGTTAATACCTTCCTACAACTGGTGAA 717
Db

RESULT 7
STRSPGIGP 1950 bp DNA linear BCT 23-DEC-1993
LOCUS Streptococcus sp. (Lancefield group G) spg gene encoding an
DEFINITION Immunoglobulin G binding protein.
ACCESSION M13825
VERSION M13825.1 GI:153822
KEYWORDS constant region; immunoglobulin binding protein.
SOURCE Streptococcus sp. (Lancefield group G; strain GX7809; DNA, clone
mGX4547.
ORGANISM Streptococcus sp.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 1950)
AUTHORS Fahnestock,S.R., Alexander,P., Nagle,J. and Filpula,D.
TITLE Gene for an immunoglobulin-binding protein from a group G
streptococcus
JOURNAL J. Bacteriol. 167, 870-880 (1986)
MEDLINE 86304178
COMMENT Draft entry and computer-readable sequence of [1] kindly provided
by S.R.Fahnestock, 07-NOV-1986. The -35 and -10 regions are
located at positions 465-470 and 487-492 respectively and an SD
sequence at 565-570.
FEATURES
source location/Qualifiers
1..1950
/organism="Streptococcus sp."
/db_xref="taxon:1306"
578..676
/note="IgG binding protein signal peptide"
578..1924
/note="IgG binding protein"
/codon_start=1
/transl_table=11
/protein_id="AAA03664.1"
/db_xref="GI:153823"
/translation="MEKEKKVKYFLRKSAPGLASVSAFLVSGTVFAVDSPIEDTPII
RNGCELTNLLGNSETFLALRNEESATADLTAAVADTVAAAEENAGAAWEAAAAAD
ALAKADALKEFNKYGVSDYYKLNINNAKTVEGIKDIAQVVSASAKKARISATDGL
SDFLKSTPAEDTVKYSIELAEAKVLANRELDKYGSDYHKNLNNAKTVEGVKVELIDE
ILAALEPKPTVKYILNGKTLKGETTTTAVDAATAEKVKFYANDNGVDGEMTVYDAAK
TFVTVEKPEVIDASELTPAVTVYKLVINGKTLKGETTTTAVDAETAERKAFKQVANDNG
VDGWMYDADKTFVTVEYVPGDAPTEPEKPEASIPVPLTPATPATKDDAKDD
TKEDAKKPEAKDDAKKAEKTLPTTGEGSNPFPTAALVAGALAVASKRKKD"
677..1921
mat_peptide
BASE COUNT 706 a 323 c 398 g 523 t
ORIGIN 1 bp upstream of HindIII site.

Alignment Scores:
Pred. No.: 5.11e-06 Length: 1950
Score: 188.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 24.20% Indels: 16
DB: 1 Gaps: 6

US-09-847-539A-6 (1-159) x STRSPGIGP (1-1950)

Qy 1 ValAspSerProIleGluGlnProArgIleProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
677 GTTGATTTCACCAATCGAAGATACCCCAATTATTGCTGATGCTGCTGAATTAACCTT 736
Db

21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Qy
|||||
737 CTGGGAATTCAGAGACAACACTGGCTTGGCTAATGAAGAGAGCTGTACAGCTGATTG 796
Db

41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
Qy
|||||
797 ACAGCAGCAGGGTAGCGGCTACTGTGGCAGCAGCGCAGCTGAAATGCTGGGGCAGCA 856
Db

61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValAl 80
Qy
|||||
857 GCTTGGGAAGCAGCGGCA---GCAGCAGATGCTCTGCAAMAGCAAGCAGATGCCCTT 913
Db

81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
Qy
|||||
914 AMCAATTCACCAATATGGAGTAAGTGACTATTACAAGAAATCTAATCAACAATGCCAA 973
Db

96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAla--- 114
Qy
|||||
974 ACTGTTGAAGGCATAAAGAGACCTTTCAAGCACAAAGCTTGTGAATCAGCGAAGAACGCGT 1033
Db

115 ---SerAspAlaTrpGlu-----LysAlaAlaThrProIleAlaLeu 127
Qy
|||||
1034 ATTTCAAGAGCAACAGATGGCTTATCTGATTTCTTGAATCGCAACACCTGCTGAAGAT 1093
Db

128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
Qy
|||||
1094 ACTGTTAAATCAATTGAATTAGCTGAAGCTAAGCTTAGCTAACACAGAA 1144
Db

RESULT 8
109115
LOCUS Sequence 23 from Patent WO 8810306.
DEFINITION I09115
ACCESSION I09115
VERSION I09115.1 GI:598187
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1950)
AUTHORS Fahnestock,S.R.
TITLE CLONED PROTEIN G VARIANT GENES AND THE PROTEIN G VARIANTS EXPRESSED
THEREFROM
JOURNAL Patent: WO 8810306-A 23 DEC-1988;
FEATURES Location/Qualifiers
source 1..1950
/organism="unknown"
BASE COUNT 706 a 323 c 398 g 523 t
ORIGIN

Alignment Scores:
Pred. No.: 5.11e-06 Length: 1950
Score: 188.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 24.20% Indels: 16
DB: 1 Gaps: 6

US-09-847-539A-6 (1-159) x I09115 (1-1950)

Qy 1 ValAspSerProIleGluGlnProArgIleProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
677 GTTGATTTCACCAATCGAAGATACCCCAATTATTGCTAATGCTGCTGAATTAACCTT 736
Db

21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Qy
|||||
737 CTGGGAATTCAGAGACAACACTGGCTTGGCTAATGAAGAGAGCTGTACAGCTGATTG 796
Db

41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
Qy
|||||

```



AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL 99269061  
MEDLINE 10336419  
PUBMED  
REFERENCE 2 (bases 1 to 469)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden  
FEATURES  
source  
1. .469  
/organism="Streptococcus pyogenes"  
/strain="AP1"  
/db\_xref="taxon:1314"  
1. .>469  
/product="GRAB"  
1. .174  
/note="Region: alpha2-macroglobulin-binding A domain"  
<1. >469  
/note="grab"  
/codon\_start=1  
/transl\_table=11  
/product="GRAB precursor"  
/protein\_id="AAD26339.1"  
/db\_xref="GI:4589081"  
/translation="VDSPIEQPRIIPNGGTLNLGNAPKIALRNEERAIDELKKQA  
IEDKEATTAEASDALEADQDALQSEAAVVKADNAASDALEALADQDALQS  
EEAVVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKERQNVNLTPTGEE"  
misc\_feature 175. .342  
/note="Region: repeat motif"  
misc\_feature 343. .469  
/note="Region: cell-wall attachment"  
BASE COUNT 178 a 90 c 100 g 101 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.39e-50 Length: 469  
Score: 734.00 Matches: 155  
Percent Similarity: 99.36% Conservative: 0  
Best Local Similarity: 99.36% Mismatches: 1  
Query Match: 97.04% Indels: 0  
DB: 1 Gaps: 0  
US-09-847-539A-6 (1-159) x AF124400 (1-469)  
QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
|||||  
Db 1 GTCGACTCACCTATCGAAGCGCTAGCAATTTATTCCTCAATGGCGAACCCTTAATTAATCTT 60  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
|||||  
Db 61 CTTGGCAATGCTCCAGAAACCTGGCATTACGTAATGAAGAAAGAGCCATTGATGAATTA 120  
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaThrAlaIleGluAlaAlaSerSerAsp 60  
|||||  
Db 121 AAAAACACAGCTATTGAGGATAAGGAGCTTACCACAGCTATAGAAGCAGCAAGTCAGAT 180  
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80  
|||||  
Db 181 GCCTTAGAGCATTAGCGGATCAACAGAGCGCTTACAAATCACAAGAGCTGGGTGTT 240  
QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
|||||  
Db 241 AAAGCGGATCAACGCTGCTAGTCAGCGCTTAGAAGCATTGGCGGATCAAAAGACGCGTTA 300  
QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTtpGluLys 120  
|||||  
Db 301 CAATCAGAGAGACTGAGTAGTTCATTCAGATACGCTGCTAGTCAGCGCTGGGAAAA 360  
QY 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140  
|||||

Db 361 GCAGCAACTCCAAFCGCTTTAGATGTTTAAAGAACTAAAGATACAAACCTGTAGTTAAA 420  
QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156  
|||||  
Db 421 AAAGAAGAAAGACAAACGTTAATACCTCTCTTACAACTGGTGAAGAG 468  
RESULT 5  
AF124403 804 bp DNA linear BCT 14-AUG-2000  
LOCUS Streptococcus pyogenes strain AP49 GRAB precursor, gene, partial  
DEFINITION cds.  
ACCESSION AF124403  
VERSION AF124403.1 GI:4589086  
KEYWORDS Streptococcus pyogenes.  
SOURCE Streptococcus pyogenes  
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 804)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL 99269061  
MEDLINE 10336419  
PUBMED  
REFERENCE 2 (bases 1 to 804)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden  
FEATURES  
Location/Qualifiers  
1. .804  
/organism="Streptococcus pyogenes"  
/strain="AP49"  
/db\_xref="taxon:1314"  
1. .>804  
/product="GRAB"  
1. .174  
/note="Region: alpha2-macroglobulin-binding A domain"  
<1. >804  
/note="grab"  
/codon\_start=1  
/transl\_table=11  
/product="GRAB precursor"  
/protein\_id="AAD26342.1"  
/db\_xref="GI:4589087"  
/translation="VDSPIEQPRIIPNGGTLNLGNAPKIALRNEERAIDELKKQA  
IEDKEATTAEASDALEADQDALQSEAAVVKADNAASDALEALADQDALQS  
EEAVVQSDNAASDALEALADQDALQSEAAVVKADNAASDALEALADQDALQSEAE  
ASVVKADNAASDALEALADQDALQSEAAVVKADNAASDALEALADQDALQSEAE  
VVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKERQNVNLTPTGEE"  
misc\_feature 175. .678  
/note="Region: repeat motif"  
misc\_feature 679. .804  
/note="Region: cell-wall attachment"  
BASE COUNT 284 a 161 c 185 g 174 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.44e-45 Length: 804  
Score: 694.00 Matches: 156  
Percent Similarity: 58.21% Conservative: 0  
Best Local Similarity: 58.21% Mismatches: 0  
Query Match: 89.32% Indels: 112  
DB: 1 Gaps: 1  
US-09-847-539A-6 (1-159) x AF124403 (1-804)  
QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
|||||  
Db 1 GTTGACTCACCTATCGAAGCGCTTCGAATTTATTCCTCAATGGCGAACCCTTAATTAATCTT 60  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
|||||

```
Pred. No.: 1.35e-50 Length: 10029
Score: 777.00 Matches: 159
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x AF000573 (1-10029)

QY 1 ValAspSerProileGluInProArgileileProAsnGlyGlyThrLeuThrAsnLeu 20
DB 6482 GTTGACTCACCTATCGACAGCGCTCGAATATTCCAAATGGCGAACCTTAACATCTT 6423
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluAlaIleAspGluLeu 40
DB 6422 CTTGGCAATGCTCCAGAAACTGGCATTAATGAAGAAGAGCCATTCATGAATTA 6363
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
DB 6362 AAAAACAAGCTATTAGAGTAAAGAAGCTACGACAGCTATAGAAGCAGCAAGTTTCAGAT 6303
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
DB 6302 GCCTTAGAAGCATTAAGCGATCAACAGACGCTTTACATATCAGAAGAGCTCGGTTGTT 6243
QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
DB 6242 AAAGCGGATACGCTGCTAGTGACGCTTAGAAGCATTTGGCGGATCAACAGACGCTTTA 6183
QY 101 GlnSerGluGluAlaGluAlaValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
DB 6182 CATTGAGAAGAGCTGAAGTACTCAATCAGATAACGCTGCTAGTGACGCTGGGAAAAA 6123
QY 121 AlaAlaThrProileAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140
DB 6122 CGAGCAACTCCAATCGCTTAGATGTAGAAAACTAAGATACAAACCTCTAGTTAA 6063
QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsnPro 159
DB 6062 AAAGAAGAAGAACAAACCTTAATACCTTCTACAACTGGTGAAGAGTCTTAACCCA 6006

RESULT 3
AF124401 468 bp DNA linear BCT 14-AUG-2000
LOCUS Streptococcus pyogenes strain KTL3 GRAB precursor, gene, partial
DEFINITION cds.
ACCESSION AF124401
VERSION AF124401.1 GI:4589082
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 468)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin
JOURNAL J. Biol. Chem. 274 (22), 15336-15344 (1999)
MEDLINE 99269061
PUBMED 10336419
REFERENCE 2 (bases 1 to 468)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES
source Location/Qualifiers
1..468
/organism="Streptococcus pyogenes"
/strain="KTL3"
/db_xref="taxon:1314"
mat_peptide 1..>468
misc_feature 1..174

CDS
/note="Region: alpha2-macroglobulin-binding A domain"
<1..>468
/feature="grab"
/codon_start=1
/transl_table=11
/product="GRAB precursor"
/protein_id="AAD26340.1"
/db_xref="gi:4589083"
/translation="VDSPIEQRIPIIPGGTTLNLGNAPKALNRBERAIDELKKQA
TEDKEATTAFAASDADALDSEEAAYVKADNAASDALALADQTDALOS
EEAEVQSDNAASDAWEKAATPIALDVKKTKDTPVVKKEERQNVNLTPTTGE"
175..342
/feature="misc_feature"
/note="Region: repeat motif"
343..468
/feature="misc_feature"
/note="Region: cell-wall attachment"
BASE COUNT 178 a 91 c 100 g 99 t
ORIGIN
Alignment Scores: 1.17e-50 Length: 468
Pred. No.: 760.00 Matches: 156
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 97.81% Gaps: 0
DB: 1
US-09-847-539A-6 (1-159) x AF124401 (1-468)
QY 1 ValAspSerProileGluInProArgileileProAsnGlyGlyThrLeuThrAsnLeu 20
DB 1 GTTGACTCACCTATCGACAGCGCTCGAATATTCCAAATGGCGAACCTTAACATCTT 60
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluAlaIleAspGluLeu 40
DB 61 CTTGGCAATGCTCCAGAAACTGGCATTAATGAAGAAGAGCCATTCATGAATTA 120
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
DB 121 AAAAACAAGCTATTGAGGATAAAGAAGCTACGACAGCTATAGAAGCAGCAAGTTTCAGAT 180
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
DB 181 GCCTTAGAAGCATTAAGCGGATCAACAGACGCTTTACATATCAGAAGAGCTCGGTTGTT 240
QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
DB 241 AAAGCGGATACGCTGCTAGTGACGCTTAGAAGCATTTGGCGGATCAACAGACGCTTTA 300
QY 101 GlnSerGluGluAlaGluAlaValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
DB 301 CAATCAGAAGAAGCTCAAGTACTTCAATCAGATAACGCTGCTAGTGACGCTGGGAAAAA 360
QY 121 AlaAlaThrProileAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140
DB 361 CGAGCAACTCCAATCGCTTAGATGTAGAAAACTAAGATACAAACCTCTAGTTAA 420
QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156
DB 421 AAAGAAGAAGAACAAACCTTAATACCTTCTACAACTGGTGAAGAG 468

RESULT 4
AF124400 469 bp DNA linear BCT 14-AUG-2000
LOCUS Streptococcus pyogenes strain AP1 GRAB precursor, gene, partial
DEFINITION cds.
ACCESSION AF124400
VERSION AF124400.1 GI:4589080
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 469)
```

```

CDS
complement(619..1911)
/gene="aroA"
/EC_number="2.5.1.19"
/notes="Best Blastp hit = sp|P43905|AROA_LACLA
3-PHOSPHOSHIMKIMATE 1-CARBOXYVINYLTRANSFERASE
(SYNTHASE) (EPSPS) >gi|1075724|pir||S52580
3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)
5-Lactococcus lactis >gi|683583|emb|CAA55180.1| (X78413)
5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus
lactis]
/codon_start=1
/transl_table=11
/product="putative 3-phosphoshikimate
1-carboxyvinyltransferase"
/protein_id="AAK34180.1"
/db_xref="GI:13622461"
/translation="MKRMKLRTNAGPLQGTIQVPGDKSISHRVILGAVAKGETRVKG
LLKGEVLTIOAFRLGVRIEKDDQLVIEGQFGLNAPCOTLNNNGSGTSMRLIA
GLLAGOPFSVKVIGDESLSKRPMDRITVPLKMGVEISGETDRQFPLOLQGNRLQP
ITYTLPISSAOVKSAALLAALQAKGTQVVEREITRNHTEMIQFGGRLIVDCKRIT
LVGQQTAGEITPGDIISSAFLWVAGLIPGSELLKNVGNVPTTGILEVVERMG
DAQELRVKRTQVVTIDILNSMGANIKATADGMIKGPVLYGANTSTYGDHRIGMM
TAIAALLVKGQVHLDKEAINTSYPTFFKDLERLCHD"
/complement(2013..2978)
/gene="Spy1353"
/complement(2013..2978)
/notes="Spy1353"
/notes="Best Blastp hit = dbj|BAB07074.1| (AP001518)
unknown conserved protein [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34181.1"
/db_xref="GI:13622462"
/translation="MAEKWFKVLKWOYEIQVFMRLQSAEMDLAIAVAYLIL
TAPFLVIAANFPYLLNIDALLRLMKQNLKPDIFRPAISAIVENIFSPSGVLGVA
TLGLTWMGRSITSLQKANKATYGASQHRDFFIGHVGLLTSLLIFLLAFALFISF
SKAAIQVLDKHLSDNITITFILLIQITVTLIFVGLMLYFLFNPKIKRITRILP
GLTFSVMTVTLNVLGVVYVVERMVDIKMFGSVNIFIMLWFLFARILILGALF
NATYQMSLGLKLEGRSGDMIAILKTLGNSDLSPSQSTEDSHTD"
/complement(2980..3840)
/gene="map"
/notes="Spy1354"
/complement(2980..3840)
/gene="map"
/EC_number="3.4.11.18"
/notes="Best Blastp hit = sp|Q92934|AMPM_BACHD METHIONINE
AMINOPEPTIDASE (MAP) (PEPTIDASE M)
>gi|11264635|pir||744405 methionyl aminopeptidase (EC
3.4.11.18) map [imported] - Bacillus halodurans
>gi|45124261|dbj|BAA75293.1| (AB017508) map homologue
(identity of 81% to B. subtilis) [Bacillus halodurans]
>gi|10172768|dbj|BAB03875.1| (AP001507) methionine
aminopeptidase [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative methionine aminopeptidase"
/protein_id="AAK34182.1"
/db_xref="GI:13622463"
/translation="MITLKAREIEAMDRAQDFLAGIHIGLRLDIIRPGVDMWEVAYV
RRCKEDNVLPQIGVDGHHMDYPATCCGLNDEVAHAPRHILKEGDLKLVDMVLS
EPDKSIVDAALDFQNVPMKKWGTSGTGGNADSCWAVGTPSDIEKOLMDVTKEA
MYRGEKAYIGNRIGDIGAQAQVESAESFGYVVRDLVGHGVGTTHIEEPMVNYGTAG
RGLRLKEGMVLVTEPMINTGTWEIDTINTGWAHKTLDGGLSCQYEHQFVITKDGPI
LTSQBERTY"
/complement(3856..5139)
/gene="Spy1355"
/complement(3856..5139)
/notes="Spy1355"
/notes="Best Blastp hit = pir||A69998 hypothetical protein
ytoi - Bacillus subtilis >gi|2293258|gb|AAC00336.1|

```

Alignment Scores:

```

(AF008220) ytoi [Bacillus subtilis]
>gi|2635392|emb|CAB14887.1| (Z99118) ytoi [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34183.1"
/db_xref="GI:13622464"
/translation="MSKHQDILYLEKLAIGKVSYSISNHLKVSOGTAYRAIKEAE
NRQIVETKPSGTVRIEKKGRVDRITDITSEIARIISDVSEVLGAGLGHGFSRPSIGA
MTQONTRRVLVKGGLITVGDRETIQILLAHENHAILVTGGFPVSKRVIEANNQRIPTV
MVTHTDTFTVATMINHALSNIRIKTDLKTVQVMIPITDGYLCESSVSEFNTLKK
TROVREPVLVDYKRKVGVSMRDVVDQLPTTKLTKVMKKNPITARPNTSLANISQKMI
FEDLNLMPVTDDEENLLGLMTRQAMENLPHQNPNTVYVSEOTLSNLTFTVYQVV
VEPTMIDSAGNSMNCVISEFLKEISIRALKTKKQKNIIIEQMMVYFLHAIQIEDELAI
YPKIITENRSNLTIDIEIPVDQVIAKAIITTKIN"
/complement(5148..5690)
/gene="Spy1356"
/complement(5148..5690)
/notes="Spy1356"
/notes="Best Blastp hit = emb|CAA03913.1| (AJ000084)
putative acetyl transferase [Proteus mirabilis]"
/codon_start=1
/transl_table=11
/product="putative acetyl transferase"
/protein_id="AAK34184.1"
/db_xref="GI:13622465"
/translation="NDIWTKLAVFAFFETPKVILRPFRYEDHWDYFSMVNDTNLYYV
FPQKTKAASDYLHVHSFIFKPGQMAIEDKATHQVIGSTRIEHYDAKTRCADIGYFL
NVAFWGQIMTEVVIVKLYLSFHEFGLKTLRIITHLENKASQKAVAKAGHQLTCFKG
SDRNTKICIKYKMYLTNDR"
/complement(5928..6581)
/gene="grab"
/notes="Spy1357"
/complement(5928..6581)
/notes="grab"
/notes="Best Blastp hit = gb|AAD26340.1|AF124401.1
(AF124401) GRAB precursor [Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="protein GRAB (protein G-related alpha 2W-binding
protein)"
/protein_id="AAK34185.1"
/db_xref="GI:13622466"
/translation="MGKEIKVKCFLLRSAPGLVAVSASVLVGSVTSVAVDSPIDPRII
PNGTTLNLGNAPKALRNEERATDELKQAEDEKATTAIEAASDALEALADGT
DALQSEEAAYVADNAASDALEALADQTDALQSEAEVQSDNAASDAWEKAATPIAL
DVKKTKDTKPVVKKEERQNVNTLPTTGEESNPFATAALAIWYSTGLVLYVSSCKEN"
/complement(6939..8198)
/gene="murZ"
/notes="Spy1358"
/complement(6939..8198)
/notes="murZ"
/notes="Best Blastp hit = sp|P19670|MUA2_BACSU
UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE 2
(ENOYLPYRUVATE TRANSFERASE) (EPT) >gi|2127262|pir||G32354
UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC
2.5.1.7) murZ - Bacillus subtilis
>gi|853767|emb|CAA89875.1| (Z49782)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis] >gi|2636235|emb|CAB15727.1| (Z99122)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis] >gi|2636247|emb|CAB15738.1| (Z99123)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="putative UDP-N-acetylglucosamine
1-carboxyvinyltransferase"
/protein_id="AAK34186.1"

```



```

ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 832)
AUTHORS Rasmussen, M., Muller, H.P. and Bjorck, L.
TITLE Protein GRAB of Streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin
JOURNAL J. Biol. Chem. 274 (22), 15336-15344 (1999)
MEDLINE 99269061
PUBMED 10336419
REFERENCE 2 (bases 1 to 832)
AUTHORS Rasmussen, M., Muller, H.P. and Bjorck, L.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES
source
1..832
/organism="Streptococcus pyogenes"
/strain="ATCC700294"
/db_xref="ATCC:700294"
/db_xref="taxon:1314"
1..99
sig_peptide
1..654
CDS
1..654
/note="grab"
/codon_start=1
/transl_table=11
/product="GRAB precursor"
/protein_id="AAD26338.1"
/db_xref="GI:4589079"
/translation="MGKEIKVCFLRSAFLVAYSALVGVSTVSAVDSPIDPRIL
PNGLTNLLAGNPEKALNRERADIDELKQAIEDKEATTAIEAASDAALEALADDT
DALQSEAAVKAADAAADALEALADTDALQSEAEVQSDNAAADAWKAATPIAL
DVKTKDTPVYKKEQRQNVNTLPTTGESNPFTTAAALIMVSTGLVTVSSKCKEN"
100..651
mat_peptide
100..651
/misc_feature
100..273
/note="Region: alpha2-macroglobulin-binding A domain"
274..441
misc_feature
/note="Region: repeat motif"
442..576
misc_feature
/note="Region: cell-wall attachment"
577..651
misc_feature
/note="Region: membrane-spanning"
BASE COUNT 304 a 143 c 169 g 216 t
ORIGIN

Alignment Scores:
Pred. No.: 1..04e-51 Length: 832
Score: 777.00 Matches: 159
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-847-539a-6 (1-159) x AF124399 (1-832)
Qy 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
Db 100 GTTGACTCACCCTATCGAAGCCTCGAATATTCCAAATGGCGGAACCTTAACCTTCT 159
|||||
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
|||||
Db 160 CTTTGCAATGCTCCAGAAAACCTGCATAGTAATGAAGAAGACCATGATGAATTA 219
|||||
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaSerSerAsp 60
|||||
Db 220 AAAAACAAGCTATTGAGATATAAGACCTAGCAGCTATAGACGACGACCAAGTTTCAGAT 279
|||||
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVal 80
|||||
Db 280 GCCTTAGAAGCATTAGCGGATCAACACAGACGCTTTTACATCAGAGAAGCTCGCTTCTT 339
|||||
Qy 81 LysAlaAspAsnAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100

```

```

|||||
Db 340 AAAGCGGATAACGCTGCTAGTGACGCTTAGAAGCATTTGGCGGATCAACAGACGCTTTA 399
|||||
Qy 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaIleSerAspAlaTrpGluLys 120
|||||
Db 400 CAATCAGAAGAGCTGAAGTAGTCTCAATCAGATAACGCTGCTAGTGACGCTGGGAAAAA 459
|||||
Qy 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140
|||||
Db 460 GCAGCACTCCAACTGCTTAGATGTTAAGAAAACATAAGATACAAAACCTGTAGTTAAA 519
|||||
Qy 141 LysGluGluArgGlnAsnValAsnThrLeuProThrGlyGluGluSerAsnPro 159
|||||
Db 520 AAGAAGAAAGACAAACAGTTAATACCTCTTCTACAACTGCTGAAGAGTCTAACCCCA 576
|||||
RESULT 2
LOCUS AE006573 10029 bp DNA linear BCT 01-JUN-2001
DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section 102 of 167 of
the complete genome.
ACCESSION AE006573 AE004092
VERSION AE006573.1 GI:13622459
KEYWORDS
SOURCE Streptococcus pyogenes M1 GAS.
ORGANISM Streptococcus pyogenes M1 GAS
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 10029)
AUTHORS Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,
Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
TITLE Complete genome sequence of an M1 strain of Streptococcus pyogenes
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
MEDLINE 21192684
PUBMED 11296296
REFERENCE 2 (bases 1 to 10029)
AUTHORS Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,
Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
TITLE Direct Submission
JOURNAL Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 SL Young Blvd,
Oklahoma City, OK 73104, USA
FEATURES
source
1..10029
/organism="Streptococcus pyogenes M1 GAS"
/strain="SF370"
/serotype="M1"
/db_xref="taxon:160490"
complement(135..626)
/feature="aok"
/feature="SPY1351"
complement(135..626)
/feature="aok"
/EC_number="2.7.1.71"
/feature="Best Blastp hit = pir1D82693 shikimate kinase
XF1335 [imported] - Xylella fastidiosa (strain 9a5c)
>gi19106332|gb|AF84144.1|AE003966_5 (AE003966) shikimate
kinase [Xylella fastidiosa]"
/codon_start=1
/transl_table=11
/product="putative shikimate kinase"
/protein_id="AAK34179.1"
/db_xref="GI:13622460"
/translation="MTKVLGFMGVGKTTVSKHLSMCKMDIAIEAKIGMSIAAFPE
OHGEIAFTIESQVLKDLFANDSIIVTGGVVVLQENRQLKRNHNLIVASFE
TLYQLKHDKKSRPLFLKYSKEAFYEQRMVFYEGSLDVRVDRHTPEVANII
EGY"
complement(619..1911)
/feature="aok"
/feature="SPY1352"

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:49:23 ; Search time 1561.09 Seconds  
(without alignments)  
2131.407 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQRIIPNGGTLTNL.....KKEERQNVNTLPTTGEENP 159

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1797636 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/uspro\_spo01/JUS09847539/runat\_10102002\_093105\_4992/app\_query.fasta\_1.526  
-DB=GenEmbl -QFAST=fastap -SUFFIX=std.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blos62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=0.5 -THR\_MIN=100 -THR\_MAX=100 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539@cgn\_1\_1\_1182 -runat\_10102002\_093105\_4992 -NCPU=3  
-NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.scs:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*

29: em.vi:\*  
30: em.htg\_hum:\*  
31: em.htg\_inv:\*  
32: em.htg\_other:\*  
33: em.htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	777	100.0	832	1	AF124399	AF124399 Streptococ
2	777	100.0	10029	1	AE006573	AE006573 Streptoco
3	760	97.8	468	1	AF124401	AF124401 Streptoco
4	754	97.0	469	1	AF124400	AF124400 Streptoco
5	694	89.3	804	1	AF124403	AF124403 Streptoco
6	692	89.1	717	1	AF124402	AF124402 Streptoco
7	188	24.2	1950	1	STRSPGIGP	M13825 Streptococ
8	188	24.2	1950	6	I09115	I09115 Sequence 23
9	187	24.1	2384	1	SGSPG	Y00428 Streptococ
10	187	24.1	2384	6	I08537	I08537 Sequence 4
11	187	24.1	2384	1	SGPROTG	X06173 Streptococ
12	186	23.9	1950	6	I08536	I08536 Sequence 2
13	154.5	19.9	1576	1	SGI48PG	X5324 Streptococ
14	151.5	19.5	1380	1	SSGMLB1	X84989 streptococ
15	139	17.9	798	6	I09107	I09107 Sequence 4
16	138.5	17.8	969	1	STRABP	M95520 Streptococ
17	131.5	16.9	3838	3	PFASA7	M10330 Plasmodium
18	128.5	16.5	13501	1	AE005258	AE005258 Escherich
19	128.5	16.5	297816	1	AP002553	AP002553 Escherich
20	122.5	15.8	3763	5	PMU19361	U19361 Petromyzon
21	119.5	15.4	9331	1	AE006458	AE006458 Escherich
22	119.5	15.4	10432	1	AE005333	AE005333 Escherich
23	119.5	15.4	222605	1	AP002555	AP002555 Escherich
24	119.5	15.4	291136	1	AP002556	AP002556 Escherich
25	119	15.3	3222	14	HFU23857	U23857 Herpesvirus
26	117	15.1	54126	3	AC006834	AC006834 Caenorhab
27	116.5	15.0	146759	10	AL604063	AL604063 Mouse DNA
28	116	14.9	35318	3	CELT03A1	AF067614 Caenorhab
29	116	14.9	44090	2	AC006784	AC006784 Caenorhab
30	116	14.9	96468	2	AC006902	AC006902 Caenorhab
31	114.5	14.7	1555	1	STRMAG	L27798 Streptococ
32	114.5	14.7	106516	1	STYPPHGM2	AL513384 Salmonell
33	114.5	14.7	210791	2	AC023354	AC023354 Mus muscu
34	114.5	14.7	222304	2	AC104834	AC104834 Mus muscu
35	114	14.7	167227	9	AC020602	AC020602 Homo sapi
36	113	14.5	146491	2	AC019950	AC019950 Drosophil
37	112.5	14.5	4469	1	SPU52008	U52008 Streptococ
38	112	14.4	1469	1	SGI48IGG	X04015 Streptococ
39	112	14.4	1469	6	AL2446	AL2446 1.5 kb Ecor
40	112	14.4	2025	5	AB025967	AB025967 Oryzias 1
41	111	14.3	1713	1	SPFCRA	X69324 S.pyogenes
42	111	14.3	2022	3	AY060997	AY060997 Drosophil
43	111	14.3	3120	1	AF324061	AF324061 Streptoco
44	111	14.3	6413	1	AF067776	AF067776 Abiotroph
45	110.5	14.2	48502	7	LAMCG	J02459 Bacterioph

# ALIGNMENTS

RESULT 1

AF124399

LOCUS

DEFINITION

complete cds.

ACCESSION

VERSION

KEYWORDS

SOURCE

AF124399

Streptococcus pyogenes strain ATCC700294

complete cds.

AF124399

AF124399.1

GI:4589078

Streptococcus pyogenes.

832 bp

DNA

linear

BCT 14-AUG-2000

GRAB precursor, gene.

THIS PAGE BLANK (USPTO)

Job time : 338.257 secs

```

MEST10-D08.T7-1 ISUM4-TN Zea mays cDNA clone MEST10-D08 5', mRNA
sequence.
ACCESSION      BG840448
VERSION        BG840448.1  GI:14206770
KEYWORDS       EST.
SOURCE         Zea mays.
ORGANISM       Zea mays
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
                clade; Panicoideae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 445)
AUTHORS        Qiu,F., Cui,F., Guo,L., Ashlock,D.A., Wen,T.J. and Schnable,P.S.
TITLE          Expressed Sequence Tags from B73 Maize Seedlings and Silks
JOURNAL        Unpublished (2001)
COMMENT        Contact: Patrick S. Schnable
                Schnable Laboratory
                Iowa State University
                G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
                Tel: 515-294-0975
                Fax: 515-294-2299
                Email: schnable@iastate.edu
PCR Primers
FORWARD: T7-1 (AA TAC GAC TCA CTA TAG)
BACKWARD: T3 (ATT AAC CCT CAC TAA AG)
Seq primer: primer T7-1 (AA TAC GAC TCA CTA TAG).

```

```

FEATURES
source
Location/Qualifiers
1. .445
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="WEST10-D08"
/clone_lib="ISUM4-TN"
/tissue_type="Seedling and silk"
/lab_host="DH10B"
/note="Vector: pT73PAC; Site 1: EcoRI; Site 2: NotI;
ds-cDNA molecules were generated as follows. First-strand
cDNA was prepared from oligo-dT selected mRNA by priming
with a NotI oligo-dT primer (5'
AACTCGAAGAATTCGGCGCGGCGGAAATTTTTTTTTTTTTT). The
resulting DNA:RNA hybrid was treated with RNase H and used
as a template for DNA PolI-catalyzed second strand
synthesis. After the addition of EcoRI adaptors, the
ds-cDNAs were digested with NotI and size-selected. The
resulting molecules were directionally cloned into the
EcoRI and NotI sites of the pT73PAC vector. The library
then went through one round of normalization to Cot value
of 5 based on the methods of Marcelo Bento Soares (Genome
Research 6: 791-806, 1996)."
131 a 99 c 120 g 95 t
BASE COUNT
ORIGIN

```

[illegible]

```

DB:          9          Gaps:          0

US-09-847-539A-6_COPY_59_86 (1-28) x A1062753 (1-666)

QY  2  AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
    |||  ::  |||  |||  |||  |||  ::  |||  |||
DB  227  GATCGCGGAATCATATAGAACACACAGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 168

QY  22  ValLysAlaAspAsnAla 27
    |||  ::  |||  |||  |||  |||
DB  167  GTACGAGGAGATCGTGCC 150

RESULT 36
LOCUS  A2699109/c
DEFINITION  A2699109 566 bp DNA linear GSS 24-JAN-2001
            RPCI-23-224J19.TV RPCI-23 Mus musculus genomic clone RPCI-23-224J19
            , DNA sequence.
ACCESSION  A2699109
VERSION    A2699109.1 GI:12416865
KEYWORDS  GSS.
SOURCE    house mouse.
          Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 666)
AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,
          ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.,
          and Fraser,C.M.
          Unpublished (1999)
TITLE     Mouse BAC End Sequences from Library RPCI-23
JOURNAL   Unpublished (1999)
COMMENT   Other_GSSs: RPCI-23-224J19.TJ
          Contact: Shaying Zhao
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: szhao@tigr.org
          Clones are derived from the mouse BAC library RPCI-23. For BAC
          library availability, please contact Pletier de Jong
          (pdejong@mail.cho.org). Clones may be purchased from BACPAC
          Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
          page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
          Plate: 224 row: J column: 19
          Seq primer: 17
          Class: BAC ends.
          Location/Qualifiers
            1..666
              /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="RPCI-23-224J19"
              /clone_lib="RPCI-23"
              /sex="Female"
              /lab_host="DH10B"
              /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
              EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
              brain genomic DNA was isolated and partially digested
              with a combination of EcoRI and EcoRI Methylase. Size
              selected DNA was cloned into the pBACe3.6 vector at the
              EcoRI sites. The ligation products were transformed into
              DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT  203 a 163 c 150 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 172 Length: 666
Score: 51.00 Matches: 11
Percent Similarity: 68.18% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 7
Query Match: 39.53% Indels: 0
DB: 12 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x A1062753 (1-666)

QY  1  SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 20
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  522  TCAGATCCCTCGGAGAGACTTAAGAAATCAGACAGATTCCTCTAAAGGCGAGAGAGAGAG 463

QY  21  ValVal 22
    |||  |||  |||  |||  |||
DB  462  GTAGTG 457

RESULT 37
LOCUS  BI637204/c
DEFINITION  BI637204 710 bp mRNA linear EST 10-SEP-2001
            SD19151.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD19151 5 similar to
            CG7289: FBan0007289 GO:() located on: 2L 22B8-22B8:: 05/19/2001,
            mRNA sequence.
ACCESSION  BI637204
VERSION    BI637204.1 GI:15539414
KEYWORDS  EST.
SOURCE    fruit fly.
          Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 710)
AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
          Lewis,S. and Rubin,G.M.
          BDGP/HIMI Drosophila EST Project
          Unpublished (2001)
          Contact: Stapleton, M.
          BDGP
          Lawrence Berkeley National Lab
          One Cyclotron Rd, Berkeley, CA 94720, USA
          Fax: 510 486 6798
          Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
          hit genomic AE003584; arm:2L [1824960,2149443]
          estimated-cyto:22B4-22D2: 05/19/2001
          Plate: SD.191 row: E column: 3
          High quality sequence stop: 578.
          Location/Qualifiers
            1..710
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"
              /clone_lib="SD Drosophila melanogaster Schneider L2 cell
              culture pOT2"
              /lab_host="DH5-alpha"
              /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
              fractionated cDNAs were directly ligated into pOT2.
              Plasmid cDNA library."
BASE COUNT  165 a 203 c 170 g 171 t
ORIGIN
Alignment Scores:
Pred. No.: 188 Length: 710
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
DB: 10 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x BI637204 (1-710)

QY  2  AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
    |||  ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB  228  GATCGCGGAATCATATAGAGACACACAGACGAGGAGGAGGAGGAGGAGGAGGAGGCGG 169

QY  22  ValLysAlaAspAsnAla 27
    |||  ::  |||  |||  |||  |||
DB  168  GTACGAGGAGATCGTGCC 151

RESULT 38

```

```

QY      '22 Vallys 23
       :
Db      5703 CTGGG 5708
       :
RESULT 18
US-08-508-004-3
: Sequence 3, Application US/08508004
: Patent No. 5582969
: GENERAL INFORMATION:
: APPLICANT: Pearson, Robert E.
: APPLICANT: Dickson, Julie A.
: APPLICANT: Hamilton, Paul T.
: APPLICANT: Little, Michael C.
: APPLICANT: Beyrer Jr., Wayne F.
: TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
: NUMBER OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
: ADDRESSEE: Company
: STREET: 1 Becton Drive
: CITY: Franklin Lakes
: STATE: NJ
: COUNTRY: US
: ZIP: 07417
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/508,004
: FILING DATE: 27-JUL-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/402,282
: FILING DATE: 10-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Fugit, Donna R.
: REGISTRATION NUMBER: 32,135
: REFERENCE/DOCKET NUMBER: P-3283
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 15664 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 222..425
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 451..747
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 747..1109
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1109..2014
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 2034..2747
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 2747..3109
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 3109..3444
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 3444..3728
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 3731..4855
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 4855..5376
: OTHER INFORMATION: /function= "potential coding
: OTHER INFORMATION: sequence"
: OTHER INFORMATION: /product= "L5 gp37 homolog"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 5382..5747
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 5837..6307
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 6403..7770
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 7770..8006
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 8033..8236
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 8244..9443
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 9450..10244
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 10371..10586
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 11115..11786
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature

```

LOCATION: 11917..12741  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 12748..14499  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 14771..15154  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 15154..15426  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 15429..15664  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
US-08-508-004-3

Alignment Scores:  
Pred. No.: 2.66e+03 Length: 15664  
Score: 42.00 Matches: 8  
Percent Similarity: 63.64% Conservative: 6  
Best Local Similarity: 36.36% Mismatches: 8  
Query Match: 32.56% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-508-004-3 (1-15664)

QY 2 AspAlaLeuGluaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluaLaVal 21  
::: ||| ::| ||||: ||| |||||

Db 5643 GAGGGCTGCCCAACATCAGGACACAGCTCGAGTGTTCGACAGACACGGCGCGG 5702

QY 22 Vallys 23

Db 5703 CTGCGG 5708

RESULT 19  
US-08-402-066-3  
Sequence 3, Application US/08402066  
Patent No. 5612182  
GENERAL INFORMATION:  
APPLICANT: Pearson, Robert E.  
APPLICANT: Dickson, Julie A.  
APPLICANT: Hamilton, Paul T.  
APPLICANT: Little, Michael C.  
APPLICANT: Beyer Jr., Wayne F.  
TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and  
ADDRESS: Company  
STREET: 1 Becton Drive  
CITY: Franklin Lakes  
STATE: NJ  
COUNTRY: US  
ZIP: 07417  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/402,066  
FILING DATE:  
CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:  
NAME: Fugit, Donna R.  
REGISTRATION NUMBER: 32,135  
REFERENCE/DOCKET NUMBER: P-3283  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15664 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 222..425  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 451..747  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 747..1109  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1109..2014  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2034..2747  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2747..3109  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3109..3444  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3444..3728  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3731..4855  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4855..5376  
OTHER INFORMATION: /function= "potential coding"  
OTHER INFORMATION: sequence"  
OTHER INFORMATION: /product= "L5 gp37 homolog"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 5382..5747  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 5837..6307  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:



```
/ NAME/KEY: misc_feature
/ LOCATION: 6403..7770
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 7770..8006
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 8033..8236
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 8244..9443
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 9450..10244
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 10371..10586
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 11115..11786
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 11917..12741
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 12748..14499
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 14771..15154
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 15154..15426
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 15429..15664
/ OTHER INFORMATION: /function= "potential open reading"
/ OTHER INFORMATION: frame"
/ US-08-402-066-3
Alignment Scores:
Pred. No.: 2.66e+03 Length: 15664
Score: 42.00 Matches: 8
Percent Similarity: 63.64% Conservative: 6
Best Local Similarity: 36.36% Mismatches: 8
Query Match: 32.56% Indels: 0
DB: 1 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-402-066-3 (1-15664)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaLaval 21
::: ||| :::: ||| |||||::||| |||||
```

```
Db 5643 GAGGCGCTGGCCACATGCAGGACAAAGCTCGACGTGTTGCAAGCAGACACGCGCGG 5702
QY 22 Vallys 23
Db 5703 CTGCGG 5708
RESULT 20
US-08-402-068-3
; Sequence 3, Application US/08402068
; Patent No. 5633159
; GENERAL INFORMATION:
; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Beyer Jr., Wayne E.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ADDRESSEE: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402,068
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 222..425
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 451..747
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747..1109
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1109..2014
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2034..2747
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
```

```

; NAME/KEY: misc_feature
; LOCATION: 2747..3109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3109..3444
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3444..3728
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3731..4855
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function= "potential coding
; OTHER INFORMATION: sequence"
; OTHER INFORMATION: /product= "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
```

```

; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; US-08-402-068-3
Alignment Scores:
Pred. No.: 2.66e+03 Length: 15664
Score: 42.00 Matches: 8
Percent Similarity: 63.64% Conservative: 6
Best Local Similarity: 36.36% Mismatches: 8
Query Match: 32.56% Indels: 0
DB: 1 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-402-068-3 (1-15664)
Qy 2 AspalalaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21
Db 5643 GAGGCGCTGGCCAACTGCAGACGCTGTTGCAAGGAGACACGCGCGCG 5702
Qy 22 Vallys 23
Db 5703 CTGCGG 5708
RESULT 21
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335.409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1
Alignment Scores:
Pred. No.: 1.65e+04 Length: 68750
Score: 42.00 Matches: 10
Percent Similarity: 59.26% Conservative: 6
Best Local Similarity: 37.04% Mismatches: 11
Query Match: 32.56% Indels: 0
DB: 3 Gaps: 0
```

```
US-09-847-539A-6_COPY_59_86 (1-28) x US-09-335-409-1 (1-68750)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
Db 53823 GATGCCCTCCGTCATGCCGCGGAGGTGCAGATCGTGGAGGCCGACGTG 53882

QY 22 VallysAlaAspAsnAlaAla 28
Db 53883 GCTCGCGCGACGATGTCGCT 53903

RESULT 22
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Alignment Scores:
Pred. No.: 1.65e+04 Length: 68750
Score: 42.00 Matches: 10
Percent Similarity: 59.26% Conservative: 6
Best Local Similarity: 37.04% Mismatches: 11
Query Match: 32.56% Indels: 0
DB: 4 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-567-969-1 (1-68750)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
Db 53823 GATGCCCTCCGTCATGCCGCGGAGGTGCAGATCGTGGAGGCCGACGTG 53882

QY 22 VallysAlaAspAsnAlaAla 28
Db 53883 GCTCGCGCGACGATGTCGCT 53903

RESULT 24
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1

Alignment Scores:
Pred. No.: 1.65e+04 Length: 68750
Score: 42.00 Matches: 10
Percent Similarity: 59.26% Conservative: 6
Best Local Similarity: 37.04% Mismatches: 11
Query Match: 32.56% Indels: 0
DB: 4 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-568-480-1 (1-68750)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
Db 53823 GATGCCCTCCGTCATGCCGCGGAGGTGCAGATCGTGGAGGCCGACGTG 53882

QY 22 VallysAlaAspAsnAlaAla 28
Db 53883 GCTCGCGCGACGATGTCGCT 53903

RESULT 25
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
```

APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,486  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 68750  
TYPE: DNA  
ORGANISM: Sorangium cellulosum  
US-09-568-486-1

Alignment Scores:  
Pred. No.: 1.65e+04 68750  
Score: 42.00 Matches: 10  
Percent Similarity: 59.26% Conservative: 6  
Best Local Similarity: 37.04% Mismatches: 11  
Query Match: 32.56% Indels: 0  
DB: 4 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-568-486-1 (1-68750)

Qy 2 AspalaleuGluAlaLeuAlaAspClnThraspalaleuGlnSerGluGluAlaAlaVal 21  
||||||| :||||| :|||||  
Db 53823 GATCCCTCCGGTCCATGCCGCGAGCTGGCGGAGGTGCAGATCGTGGAGCGCGACGTG 53882

Qy 22 VallysAlaAspAsnAlaAla 28  
: : : : :  
Db 53883 GCTCGGCCGCGAGATGCGCT 53903

## RESULT 26

US-09-568-472-1  
Sequence 1, Application US/09568472  
Patent No. 6358719

GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,472  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 68750  
TYPE: DNA  
ORGANISM: Sorangium cellulosum  
US-09-568-472-1

Alignment Scores:  
Pred. No.: 1.65e+04 68750  
Score: 42.00 Matches: 10  
Percent Similarity: 59.26% Conservative: 6  
Best Local Similarity: 37.04% Mismatches: 11  
Query Match: 32.56% Indels: 0  
DB: 4 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-568-472-1 (1-68750)

Qy 2 AspalaleuGluAlaLeuAlaAspClnThraspalaleuGlnSerGluGluAlaAlaVal 21  
||||||| :||||| :|||||  
Db 53823 GATCCCTCCGGTCCATGCCGCGAGCTGGCGGAGGTGCAGATCGTGGAGCGCGACGTG 53882

Db 53823 GATCCCTCCGGTCCATGCCGCGAGCTGGCGGAGGTGCAGATCGTGGAGCGCGACGTG 53882  
Qy 22 VallysAlaAspAsnAlaAla 28  
: : : : :  
Db 53883 GCTCGGCCGCGAGATGCGCT 53903

## RESULT 27

PCT-US93-01676A-3  
Sequence 3, Application PC/TUS9301676A  
GENERAL INFORMATION:  
APPLICANT: Stambrook, Peter J.  
TITLE OF INVENTION: MUTAGENICITY TESTING USING REPORTER  
GENES WITH MODIFIED METHYLATION FREQUENCIES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01676A  
FILING DATE: 19930226

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/842,644  
FILING DATE: 02-FEB-1992

ATTORNEY/AGENT INFORMATION:  
NAME: Cooper, Iver P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: STAMBROOK 1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1129 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:

NAME/KEY: CDS  
LOCATION: 10..1122  
PCT-US93-01676A-3

Alignment Scores:  
Pred. No.: 154 Length: 1129  
Score: 41.00 Matches: 10  
Percent Similarity: 65.22% Conservative: 5  
Best Local Similarity: 43.48% Mismatches: 8  
Query Match: 31.78% Indels: 0  
DB: 5 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x PCT-US93-01676A-3 (1-1129)

Qy 1 SerAspalaleuGluAlaLeuAlaAspClnThraspalaleuGlnSerGluGluAlaAla 20  
: : : : :  
Db 1036 GCTGATTCACATGCAGCTACGAAGACAGGTTTCCAGACTGGAAGTGGCGGACGCT 1095

Qy 21 ValVallys 23  
: : :  
Db 1096 CTGCCCAAG 1104

## RESULT 28

US-07-876-280-8

```
; Sequence 8, Application US/07876280
; Patent No. 5262158
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.
; APPLICANT: Cannon, Raymond J.C.
; APPLICANT: Bagley, Angela L.
; TITLE OF INVENTION: No. 5262158el Bacillus thuringiensis Isolates for
; TITLE OF INVENTION: Controlling Acarides
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/876,280
; FILING DATE: 19920430
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: M/S 104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1185 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; INDIVIDUAL ISOLATE: PS69D1
; IMMEDIATE SOURCE:
; CLONE: E. COLI NM522(pMYC2317) NRRL B-18816
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..1185
US-07-876-280-8
Alignment Scores:
Pred. No.: 163 Length: 1185
Score: 41.00 Matches: 9
Percent Similarity: 59.09% Conservative: 4
Best Local Similarity: 40.91% Mismatches: 9
Query Match: 31.78% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6_COPY_59_86 (1-28) x US-07-876-280-8 (1-1185)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValVallysAla 24
Db 691 GAAGACTTACTAAGAAAGACTAGACGACCTGAAAAAAGAGCTTGGAAAGCTGCTATTAAAGCA 750
QY 25 AspAsn 26
Db 751 GAGAAAT 756
RESULT 29
US-08-049-783-5
; Sequence 5, Application US/08049783
; Patent No. 5439881
; GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E
; APPLICANT: Schwab, George E
; APPLICANT: Payne, Jewel M
; TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active
; TITLE OF INVENTION: Toxins Cloned from Bacillus thuringiensis Isolates
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,783
; FILING DATE: 19930419
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 5 (PS69D1):
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1185 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; INDIVIDUAL ISOLATE: PS69D1
; IMMEDIATE SOURCE:
; LIBRARY: LAMBDA GEM(TW)-11 LIBRARY OF KENNETH NARVA
; CLONE: PS69D1A
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..1185
US-08-049-783-5
Alignment Scores:
Pred. No.: 163 Length: 1185
Score: 41.00 Matches: 9
Percent Similarity: 59.09% Conservative: 4
Best Local Similarity: 40.91% Mismatches: 9
Query Match: 31.78% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6_COPY_59_86 (1-28) x US-08-049-783-5 (1-1185)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValVallysAla 24
Db 691 GAAGACTTACTAAGAAAGACTAGACGACCTGAAAAAAGAGCTTGGAAAGCTGCTATTAAAGCA 750
QY 25 AspAsn 26
Db 751 GAGAAAT 756
RESULT 30
US-08-316-301A-9
; Sequence 9, Application US/08316301A
; Patent No. 5753492
; GENERAL INFORMATION:
; APPLICANT: Schnepf, Harry E.
; APPLICANT: Schwab, George E.
```



```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/222,594
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/904,278
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sanders, Jay M.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: MA-702
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1185 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS69D1
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC2317) NRRL B-18816
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1..1185
US-09-222-594-9

Alignment Scores:
Pred. No.: 163 Length: 1185
Score: 41.00 Matches: 9
Percent Similarity: 59.09% Conservative: 4
Best Local Similarity: 40.91% Mismatches: 9
Query Match: 31.78% Indels: 0
DB: Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-222-594-9 (1-1185)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLaValValLysAla 24
||| ||| :: ||| |||:: ||| ||| ::||| |||
Db 691 GAGAGGTTACTAAGAAGTAGACGACCTGAAAAAGAGTGTGGAAGCTGCTATTAAGCA 750

Qy 25 AspAsn 26
Db 751 GAGAA 756

RESULT 34
PCT-US92-03624-9
Sequence 9, Application PC/TUS9203624
GENERAL INFORMATION:
APPLICANT: Schnepf, Harry E.
APPLICANT: Schwab, George E.
APPLICANT: Payne, Jewel M.
APPLICANT: Narva, Kenneth E.
APPLICANT: Foncecrada, Luis
TITLE OF INVENTION: Novel Nematode-Active Toxins and Genes
TITLE OF INVENTION: Which Code Therefor
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA

```





REGISTRATION NUMBER: 35,303  
REFERENCE/DOCKET NUMBER: 93,509  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
TELEX: 910-221-5317  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..30  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31..1626  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 1626..1680  
US-08-140-729A-2

Alignment Scores:  
Pred. No.: 252 Length: 1680  
Score: 41.00 Matches: 10  
Percent Similarity: 56.52% Conservative: 3  
Best Local Similarity: 43.48% Mismatches: 10  
Query Match: 31.78% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-140-729A-2 (1-1680)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAla 24  
||||| ||||||| ||||| ||||||| :  
Db 399 GAGGCCAAAGTAGGCACACGCGATGCCGCCAGAGCCGCGAGCGAGCTGGCATCGAGCGA 340  
QY 25 AspAsnAla 27  
|||  
Db 339 GCGGCGGCC 331

RESULT 37  
US-08-546-666-2/c  
Sequence 2, Application US/08546666  
Patent No. 5776774  
GENERAL INFORMATION:  
APPLICANT: Amara, Susan G  
APPLICANT: Arriza, Jeffrey L  
TITLE OF INVENTION: Amino Acid Transporters and Uses  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allegretti & Witcoff, Ltd.  
STREET: 10 South Wacker Drive, Suite 3000  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/546,666  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/140,729  
FILING DATE: 20 OCT 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: NO. 5776774nan, Kevin E  
REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 93,509  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
TELEX: 910-221-5317  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..30  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31..1626  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 1626..1680  
US-08-546-666-2

Alignment Scores:  
Pred. No.: 252 Length: 1680  
Score: 41.00 Matches: 10  
Percent Similarity: 56.52% Conservative: 3  
Best Local Similarity: 43.48% Mismatches: 10  
Query Match: 31.78% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-546-666-2 (1-1680)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAla 24  
||||| ||||||| ||||| ||||||| :  
Db 399 GAGGCCAAAGTAGGCACACGCGATGCCGCCAGAGCCGCGAGCGAGCTGGCATCGAGCGA 340  
QY 25 AspAsnAla 27  
|||  
Db 339 GCGGCGGCC 331

RESULT 38  
US-08-916-745-2/c  
Sequence 2, Application US/08916745  
Patent No. 5840516  
GENERAL INFORMATION:  
APPLICANT: Amara, Susan G  
APPLICANT: Arriza, Jeffrey L  
TITLE OF INVENTION: Amino Acid Transporters and Uses  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allegretti & Witcoff, Ltd.  
STREET: 10 South Wacker Drive, Suite 3000  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/916,745  
FILING DATE: 19-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/140,729  
FILING DATE: 20-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: NO. 5840516nan, Kevin E  
REGISTRATION NUMBER: 35,303  
REFERENCE/DOCKET NUMBER: 93,509

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
TELEX: 910-221-5317  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..30  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31..1626  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 1626..1680  
US-08-916-745-2

Alignment Scores:  
Pred. No.: 252 Length: 1680  
Score: 41.00 Matches: 10  
Percent Similarity: 56.52% Conservative: 3  
Best Local Similarity: 43.48% Mismatches: 10  
Query Match: 31.78% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-916-745-2 (1-1680)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24

Db 399 GAGGCCAAAGTAGCGGACACGGATGCCCGACAGAGCCCGAGGAGCGCTGGCATCGAGCGA 340

Qy 25 AspAsnAla 27

Db 339 GCGCGCGCC 331

## RESULT 39

US-09-042-929-2/c

Sequence 2, Application US/09042929

Patent No. 5919628

GENERAL INFORMATION:

APPLICANT: Amara, Susan G

APPLICANT: Arriza, Jeffrey L

TITLE OF INVENTION: Amino Acid Transporters and Uses

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Allegretti &amp; Witcoff, Ltd.

STREET: 10 South Wacker Drive, Suite 3000

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/042,929

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/140,729

FILING DATE: 20 OCT 1993

ATTORNEY/AGENT INFORMATION:

NAME: No. 5919628nan, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 93,509

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234

TELEX: 910-221-5317

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1680 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: 5'UTR

LOCATION: 1..30

FEATURE:

NAME/KEY: CDS

LOCATION: 31..1626

FEATURE:

NAME/KEY: 3'UTR

LOCATION: 1626..1680  
US-09-042-929-2

Alignment Scores:  
Pred. No.: 252 Length: 1680  
Score: 41.00 Matches: 10  
Percent Similarity: 56.52% Conservative: 3  
Best Local Similarity: 43.48% Mismatches: 10  
Query Match: 31.78% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-042-929-2 (1-1680)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24

Db 399 GAGGCCAAAGTAGCGGACACGGATGCCCGACAGAGCCCGAGGAGCGCTGGCATCGAGCGA 340

Qy 25 AspAsnAla 27

Db 339 GCGCGCGCC 331

## RESULT 40

US-08-546-661-2/c

Sequence 2, Application US/08546661

Patent No. 5919699

GENERAL INFORMATION:

APPLICANT: Amara, Susan G

APPLICANT: Arriza, Jeffrey L

TITLE OF INVENTION: Amino Acid Transporters and Uses

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Allegretti &amp; Witcoff, Ltd.

STREET: 10 South Wacker Drive, Suite 3000

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/546,661

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/140,729

FILING DATE: 20 OCT 1993

ATTORNEY/AGENT INFORMATION:

NAME: No. 5919699nan, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 93,509

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000

```

; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..30
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..1626
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 1626..1680
; US-08-546-661-2

```

```

Alignment Scores:
Pred. No.: 252      Length: 1680
Score: 41.00        Matches: 10
Percent Similarity: 56.52% Conservative: 3
Best Local Similarity: 43.48% Mismatches: 10
Query Match: 31.78% Indels: 0
DB: 2 Gaps: 0

```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-546-661-2 (1-1680)

```

OY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAla 24
   |||||
Db 399 GAGGCCAAAGTAGCGGACACGGATGCCGCCAGACGCCCGAGGCGAGTGGCATCGAGCGA 340
   |||||
OY 25 AspAsnAla 27
   |||
Db 339 GCGGGCGCC 331

```

Search completed: October 13, 2002, 03:45:56  
Job time : 674.134 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 00:48:06 ; Search time 11.3797 seconds  
(without alignments)  
273.300 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALEALADQTDALQSEEAHVVKADNAA 28

Scoring table: BLOSUM62

Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	129	100.0	28	21	AA1980.DAT
2	129	100.0	141	21	AA1981.DAT
3	129	100.0	155	21	AA1982.DAT
4	129	100.0	159	21	AA1983.DAT
5	129	100.0	167	21	AA1984.DAT
6	129	100.0	184	21	AA1985.DAT
7	129	100.0	217	21	AA1986.DAT
8	129	100.0	259	21	AA1987.DAT
9	129	100.0	271	21	AA1988.DAT
10	57	44.2	192	18	AAW20445
11	57	44.2	222	18	AAW20638

12	57	44.2	412	19	AA10994
13	50	38.8	1164	22	ABG30268
14	47	36.4	258	20	AA136771
15	46	35.7	705	22	AA190301
16	45	34.9	502	21	AA154580
17	45	34.9	1362	20	AA199481
18	45	34.9	1447	20	AA181029
19	45	34.9	1637	22	ABG60342
20	44	34.1	42	22	ABG01077
21	44	34.1	274	21	AA158293
22	44	34.1	274	21	AA181028
23	44	34.1	284	22	AA178512
24	44	34.1	296	21	AA158292
25	44	34.1	296	21	AA181027
26	44	34.1	344	19	AA161234
27	44	34.1	358	21	AA158291
28	44	34.1	358	21	AA181026
29	44	34.1	371	21	AA181672
30	44	34.1	421	22	AB171817
31	44	34.1	429	22	AB165976
32	44	34.1	514	19	AA180400
33	44	34.1	592	22	AA192957
34	44	34.1	951	20	AA181028
35	44	34.1	1412	22	AA159278
36	44	34.1	1420	20	AA181025
37	44	34.1	1522	20	AA121975
38	44	34.1	1630	22	AB159874
39	43	33.3	33	21	AA195962
40	43	33.3	92	21	AA147712
41	43	33.3	226	21	AA121230
42	43	33.3	254	22	AA12557
43	43	33.3	258	20	AA135145
44	43	33.3	407	20	AA105072
45	43	33.3	418	20	AA105071

#### ALIGNMENTS

RESULT: 1

AA171038  
ID AA171038 standard; peptide: 28 AA.

AC AA171038:

DT 29-AUG-2000 (first entry)

DE Streptococcus pyogenes strain SF370 GRAB protein fragment #3.

XX GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

XX immune response; Streptococcus pyogenes infection.

OS Streptococcus pyogenes.

XX Streptococcus pyogenes.

PN WO200026240-A2.

PD 11-MAY-2000.

XX 02-NOV-1999; 99WO-GB03631.

XX 02-NOV-1998; 98GB-0023975.

XX (ACTI-) ACTINOVA LTD.

XX Bjorck LH, Rasmussen M;

XX WPI; 2000-365572/31.

PT New alpha2M binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein -

XX Claim 3; Page 56; 67pp; English.

PS



```
RESULT 4
AAV71041
ID AAV71041 standard; peptide; 159 AA.
XX AC
XX AAV71041;
DT 29-AUG-2000 (first entry)
XX STreptococcus pyogenes strain SF370 GRAB protein fragment #5.
DE STreptococcus pyogenes strain SF370 GRAB protein fragment #5.
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX STreptococcus pyogenes.
OS STreptococcus pyogenes.
XX WO200026240-A2.
XX 11-MAY-2000.
XX 02-NOV-1999; 99WO-GB03631.
XX 02-NOV-1998; 98GB-0023975.
XX (ACTI-) ACTINOVA LTD.
PI Bjorck LH, Rasmussen M;
XX WPT; 2000-365572/31.
XX N-PSDB; AAD00564.
XX New alpha2M binding protein for generating a protective immune response
to group A streptococcus and purifying the binding protein -
XX Claim 5; Page 57-58; 67pp; English.
XX The patent discloses a new family of proteins termed GRAB (protein G
related alpha2M binding protein) from Streptococcus pyogenes which have
the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
protein G of group G Streptococcus. GRAB protein and peptides derived
from it are used in vaccine compositions for generating a protective
immune response against group A Streptococcus. Antibodies against GRAB
are useful for treating Streptococcus pyogenes infections. The protein
is also useful for purifying alpha2M from a sample. The present sequence
is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding
to residues 34-192 and devoid of the membrane spanning region. This
fragment is useful in vaccine composition.
XX SQ Sequence 159 AA;
Query Match 100.0%; Score 129; DB 21; Length 159;
Best Local Similarity 100.0%; Pred. No. 3.4e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SDALALADQTDALQSEEAAYVKADNAA 28
Db 59 SDALALADQTDALQSEEAAYVKADNAA 86
RESULT 5
AAV71046
ID AAV71046 standard; Protein; 167 AA.
XX AC
XX AAV71046;
DT 29-AUG-2000 (first entry)
XX STreptococcus pyogenes strain KTL3 partial GRAB protein.
DE STreptococcus pyogenes strain KTL3 partial GRAB protein.
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX
```

```
OS STreptococcus pyogenes.
XX WO200026240-A2.
XX 11-MAY-2000.
XX 02-NOV-1999; 99WO-GB03631.
XX 02-NOV-1998; 98GB-0023975.
XX (ACTI-) ACTINOVA LTD.
PI Bjorck LH, Rasmussen M;
XX WPT; 2000-365572/31.
XX N-PSDB; AAD00564.
XX New alpha2M binding protein for generating a protective immune response
to group A streptococcus and purifying the binding protein -
XX Claim 5; Page 62-63; 67pp; English.
XX The patent discloses a new family of proteins termed GRAB (protein G
related alpha2M binding protein) from Streptococcus pyogenes which have
the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
protein G of group G Streptococcus. GRAB protein and peptides derived
from it are used in vaccine compositions for generating a protective
immune response against group A Streptococcus. Antibodies against GRAB
are useful for treating Streptococcus pyogenes infections. The protein
is also useful for purifying alpha2M from a sample. The present sequence
is a partial GRAB protein from S. pyogenes strain KTL3.
XX The protein has alpha2M binding region and is useful in vaccine
composition.
XX SQ Sequence 167 AA;
Query Match 100.0%; Score 129; DB 21; Length 167;
Best Local Similarity 100.0%; Pred. No. 3.6e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SDALALADQTDALQSEEAAYVKADNAA 28
Db 70 SDALALADQTDALQSEEAAYVKADNAA 97
RESULT 6
AAV71039
ID AAV71039 standard; Protein; 184 AA.
XX AC
XX AAV71039;
DT 29-AUG-2000 (first entry)
XX STreptococcus pyogenes strain SF370 mature GRAB protein.
DE STreptococcus pyogenes strain SF370 mature GRAB protein.
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX STreptococcus pyogenes.
XX OS
XX Key Location/Qualifiers
FH Binding-site 1..58
FT /label= alpha2-macroglobulin_binding_site
FT Region 59..86
FT /label= R1
FT /note= "repeat region"
FT Region 87..114
FT /label= R2
FT /note= "repeat region"
FT Region 115..159
FT /label= Cell_wall_spanning_region
FT Region 150..155
```

```

FT FT /note= "consensus sequence for gram-positive
FT FT surface cell wall anchored proteins"
FT FT 160..184
FT FT /label= Membrane_spanning_region
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACTI-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
XX
XX New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein
XX
XX Claim 5; Page 56; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G Streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A Streptococcus. Antibodies against GRAB
XX are useful for treating Streptococcus pyogenes infections. The protein
XX is also useful for purifying alpha2M from a sample. The present sequence
XX is a mature GRAB protein from S. pyogenes strain SF370 without the
XX signal sequence. This sequence is capable of binding alpha2M
XX and useful in vaccine composition.
XX
XX SQ Sequence 184 AA;
XX
XX Query Match 100.0%; Score 129; DB 21; Length 184;
XX Best Local Similarity 100.0%; Pred. No. 4e-11;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SDALEALADQTDALQSEEAHVVKADNAA 28
XX |
XX Db 59 SDALEALADQTDALQSEEAHVVKADNAA 86
XX
XX RESULT 7
XX AAY71042
XX ID AAY71042 standard; Protein; 217 AA.
XX
XX AC AAY71042;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX XX Streptococcus pyogenes strain SF370 full-length GRAB protein.
XX
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX immune response; Streptococcus pyogenes infection.
XX
XX OS Streptococcus pyogenes.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..33
XX FT /label= Signal_sequence
XX FT Protein 34..217
XX FT /label= Mature_GRAB_protein
XX FT Binding-site 34..91
XX FT /label= alpha2M_binding_site
XX FT Domain 34..68
XX FT /note= "shows homology to E domain of protein G"
XX FT Region 92..119

```

```

FT Region /label= Repeat_region_1
FT 120..147
FT /label= Repeat_region_2
FT 148..192
FT /label= Cell_wall_spanning_region
FT 183..188
FT /note= "consensus sequence for gram-positive
FT surface cell wall anchored proteins"
FT 193..217
FT /label= Membrane_spanning_region
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACTI-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
XX N-PSDB; AAD00559, AAD00560.
XX
XX New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein
XX
XX Claim 5; Fig 2B; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G Streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A Streptococcus. Antibodies against GRAB
XX are also useful for purifying alpha2M from a sample. The present sequence
XX is a full-length GRAB protein from S. pyogenes strain SF370.
XX
XX SQ Sequence 217 AA;
XX
XX Query Match 100.0%; Score 129; DB 21; Length 217;
XX Best Local Similarity 100.0%; Pred. No. 4.9e-11;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SDALEALADQTDALQSEEAHVVKADNAA 28
XX |
XX Db 92 SDALEALADQTDALQSEEAHVVKADNAA 119
XX
XX RESULT 8
XX AAY71043
XX ID AAY71043 standard; Protein; 259 AA.
XX
XX AC AAY71043;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX XX Streptococcus pyogenes strain KTL9 partial GRAB protein.
XX
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX immune response; Streptococcus pyogenes infection.
XX
XX OS Streptococcus pyogenes.
XX
XX PN WO200026240-A2.
XX
XX PD 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.

```



XX 02-NOV-1998; 98GB-0023975.  
 PR (ACTI-) ACTINOVA LTD.  
 PA  
 XX Bjorck LH, Rasmussen M;  
 PI  
 XX WPI; 2000-365572/31.  
 DR N-PSDB; AAD00561.  
 DR  
 XX New alpha2M binding protein for generating a protective immune response  
 PT to group A streptococcus and purifying the binding protein -  
 PT  
 XX Claim 5; Page 59-60; 67pp; English.  
 PS  
 XX The patent discloses a new family of proteins termed GRAB (protein G  
 CC related alpha2M binding protein) from Streptococcus pyogenes which have  
 CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
 CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2M from a sample. The present sequence  
 CC is a partial GRAB protein from S. pyogenes strain KTL9.  
 CC The protein has alpha2M binding region and is useful in vaccine  
 CC composition.  
 CC  
 XX Sequence 259 AA;  
 SQ  
 Query Match 100.0%; Score 129; DB 21; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 6e-11;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SDALEALADQTDALQSEEAHVVKADNAA 28  
 DB 107 SDALEALADQTDALQSEEAHVVKADNAA 134  
 RESULT 9  
 AAY71045  
 ID AAY71045 standard; Protein; 271 AA.  
 XX  
 AC AAY71045;  
 XX  
 DT 29-AUG-2000 (first entry)  
 DE Streptococcus pyogenes strain AP49 partial GRAB protein.  
 XX  
 KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
 KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
 KW immune response; Streptococcus pyogenes infection.  
 XX  
 OS Streptococcus pyogenes.  
 XX  
 PN WO200026240-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 02-NOV-1999; 99WO-GB03631.  
 XX  
 PR 02-NOV-1998; 98GB-0023975.  
 XX  
 XX (ACTI-) ACTINOVA LTD.  
 PA  
 PI Bjorck LH, Rasmussen M;  
 XX  
 XX WPI; 2000-365572/31.  
 DR N-PSDB; AAD00563.  
 DR  
 XX New alpha2M binding protein for generating a protective immune response  
 PT to group A streptococcus and purifying the binding protein -  
 PT  
 XX Claim 5; Page 61-62; 67pp; English.  
 PS

XX The patent discloses a new family of proteins termed GRAB (protein G  
 CC related alpha2M binding protein) from Streptococcus pyogenes which have  
 CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
 CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2M from a sample. The present sequence  
 CC is a partial GRAB protein from S. pyogenes strain AP49.  
 CC The protein has alpha2M binding region and is useful in vaccine  
 CC composition.  
 CC  
 XX Sequence 271 AA;  
 SQ  
 Query Match 100.0%; Score 129; DB 21; Length 271;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-11;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SDALEALADQTDALQSEEAHVVKADNAA 28  
 DB 174 SDALEALADQTDALQSEEAHVVKADNAA 201  
 RESULT 10  
 AAW20445  
 ID AAW20445 standard; Protein; 192 AA.  
 XX  
 AC AAW20445;  
 XX  
 DT 11-JUL-1997 (first entry)  
 DE H. pylori secreted or periplasmic protein, 35336707.aa.  
 XX  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacterium; life cycle; activator;  
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 KW diagnosis.  
 XX  
 OS Helicobacter pylori.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 181 /note= "encoded by GYG"  
 FT Misc-difference 183 /note= "encoded by GYC"  
 FT Misc-difference 187 /note= "encoded by AWT"  
 FT Misc-difference 192 /note= "encoded by AWT"  
 XX  
 PN WO9640893-A1.  
 XX  
 PD 19-DEC-1996.  
 XX  
 PF 06-JUN-1996; 96WO-US09122.  
 XX  
 PR 01-APR-1996; 96US-0630405.  
 PR 07-JUN-1995; 95US-0487032.  
 XX  
 XX (ASTR ) ASTRA AB.  
 XX  
 PI Berglindh OT, Smith D, Mellgaard BL;  
 XX  
 DR WPI; 1997-052306/05.  
 DR N-PSDB; AAT67618.  
 XX  
 PT Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 PT infection, and to detect Helicobacter  
 XX  
 PS Claim 72; Page 619; 1481pp; English.  
 XX

CC This sequence represents a *H. pylori* secreted or periplasmic protein.  
 CC The protein may be used in a vaccine to prevent or treat *H. pylori*  
 CC infection or to identify *H. pylori* polypeptide binding compounds,  
 CC useful as potential *H. pylori* life cycle activators or inhibitors.  
 CC The genomic sequence of *H. pylori* (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely *H. pylori* antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide  
 CC production, e.g. in *E. coli* hosts.

XX  
 SQ Sequence 192 AA;

Query Match 44.2%; Score 57; DB 18; Length 192;  
 Best Local Similarity 52.2%; Pred. No. 1.3;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 4 LEALADOTDALQSEAAVVKADN 26  
 :||:| ||| | | | | | | | |  
 Db 49 IQALQEQIDALDSQEKVVKWDN 71

## RESULT 11

AAW20638  
 ID AAW20638 standard; protein; 222 AA.

XX  
 AC AAW20638;

XX  
 DT 14-JUL-1997 (first entry)

XX  
 DE *H. pylori* secreted or periplasmic protein, 02cel0216orf1.

XX  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.  
 XX  
 OS Helicobacter pylori.

XX  
 PN W09640893-AL.

XX  
 PD 19-DEC-1996.

XX  
 PF 06-JUN-1996; 96WO-US09122.

XX  
 PR 01-APR-1996; 96US-0630405.

XX  
 PR 07-JUN-1995; 95US-0487032.

XX  
 PA (ASTR ) ASTRA AB.

XX  
 PI Berglindh OT, Smith D, Mellgaerd BL;

XX  
 DR WPI; 1997-052306/05.

XX  
 DR N-PSDB; AAT67891.

XX  
 PT Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent *H. pylori*  
 PT infection, and to detect Helicobacter

XX  
 PS Claim 72; Page 1062; 1481pp; English.

XX  
 CC The present sequence is a *H. pylori* secreted or periplasmic protein.  
 CC The protein may be used in a vaccine to prevent or treat *H. pylori*  
 CC infection or to identify *H. pylori* polypeptide binding compounds,  
 CC useful as potential *H. pylori* life cycle activators or inhibitors.  
 CC The genomic sequence of *H. pylori* (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely *H. pylori* antigens for vaccine development, the amino

CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide  
 CC production, e.g. in *E. coli* hosts.

XX  
 SQ Sequence 222 AA;

Query Match 44.2%; Score 57; DB 18; Length 222;  
 Best Local Similarity 52.2%; Pred. No. 1.5;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 4 LEALADOTDALQSEAAVVKADN 26  
 :||:| ||| | | | | | | | |  
 Db 77 IQALQEQIDALDSQEKVVKWDN 99

## RESULT 12

AAV10994  
 ID AAV10994 standard; protein; 412 AA.

XX  
 AC AAV10994;

XX  
 DT 08-JUN-1999 (first entry)

XX  
 DE *H. pylori* ORF 09cel0413\_35336707\_f2\_9 secreted protein.

XX  
 KW vaccine; probe; diagnostic; ORF; cell envelope protein;

XX  
 KW secreted protein; cellular protein.

XX  
 OS Helicobacter pylori.

XX  
 PN W09818323-AL.

XX  
 PD 07-MAY-1998.

XX  
 PF 28-OCT-1997; 97WO-US19575.

XX  
 PR 14-JUL-1997; 97US-0891928.

XX  
 PR 28-OCT-1996; 96US-0739150.

XX  
 PR 06-DEC-1996; 96US-0759739.

XX  
 PA (ASTR ) ASTRA AB.

XX  
 PI Alm RA, Smith D;

XX  
 DR WPI; 1998-271811/24.

XX  
 DR N-PSDB; AAX30461.

XX  
 PT Helicobacter pylori nucleic acids and proteins - used to develop  
 PT products for the detection, prevention and treatment of *H. pylori*  
 PT infections

XX  
 PS Claims 27, 31; Page 201-202; 279pp; English.

XX  
 CC Recombinant or substantially pure preparations of *H. pylori* polypeptides  
 CC are disclosed, together with the nucleic acids encoding them. In all,  
 CC 73 ORFs are shown. The proteins are variously cell envelope proteins,  
 CC secreted proteins or other cellular proteins. Vaccines containing the  
 CC nucleic acids or proteins are claimed, as are probes containing at least  
 CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful  
 CC for treating or reducing the risk of *H. pylori* infections, and the  
 CC probes can be used diagnostically for detecting the presence of  
 CC Helicobacter in a sample. The products are also of use in screening  
 CC for compounds having the ability to interfere with the *H. pylori* life  
 CC cycle or to inhibit *H. pylori* infection.

XX  
 SQ Sequence 412 AA;

Query Match 44.2%; Score 57; DB 19; Length 412;  
 Best Local Similarity 52.2%; Pred. No. 3.1;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 4 LEALADQTDALQSEEAAYVKADN 26  
::|| :||| :||| :||| :|||  
Db 49 IQALQEQIDALDQSEKVKWSDN 71

RESULT 13  
ABC30268  
ID ABG30268 standard; Protein; 1164 AA.

XX AC ABG30268;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #30259.

XX KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS94455.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX PS Claim 20; SEQ ID No 60627; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABC00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1164 AA;

Query Match 38.8%; Score 50; DB 22; Length 1164;  
Best Local Similarity 45.8%; Pred. No. 1.1e+02;

Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 3 ALEALADQTDALQSEEAAYVKADN 26  
:| :||| :||| :||| :|||

Db 581 AVIQLTPQTEAVRSEDAPVAPRDN 604

RESULT 14

AAV36771  
ID AAV36771 standard; Protein; 258 AA.

XX AC AAV36771;

XX DT 07-OCT-1999 (first entry)

XX DE Chlamydia trachomatis cellular envelope protein.

XX KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
XX KW paratrachoma; inclusion conjunctivitis; genital disease; peritheatitis;  
XX KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
XX KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX OS Chlamydia trachomatis.

XX PN WO9928475-A2.

XX PD 10-JUN-1999.

XX PF 27-NOV-1998; 98WO-IB01939.

XX PR 04-NOV-1998; 98US-0107077.

XX PR 28-NOV-1997; 97FR-0015041.

XX PR 17-DEC-1997; 97FR-0016034.

XX PA (GEST ) GENSET.

XX PI Griffais R;

XX DR WPI; 1999-371125/31.

XX PT Genome sequence of Chlamydia trachomatis

XX PS Disclosure; Page 668-669; 1755pp; English.

XX CC AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as  
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
CC can also be used to control growth of the microorganism. Chlamydia  
CC trachomatis is responsible for a large number of diseases, e.g. eye  
CC diseases such as conventional trachoma, nonendemic trachoma,  
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
CC peritheatitis, bartholinitis; pneumopathy in breast feeding infants;  
CC and venereal lymphogranulomatosis. The polypeptides of the invention  
CC may be of use in treating these diseases.

XX SQ Sequence 258 AA;

Query Match 36.4%; Score 47; DB 20; Length 258;

Best Local Similarity 36.4%; Pred. No. 51;  
Matches 8; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 4 LEALADQTDALQSEEAAYVKAD 25

:: :||| :||| :||| :|||

Db 71 IQEISDQINKLENQQAAYVKMD 92

RESULT 15

AAG90301  
ID AAG90301 standard; Protein; 705 AA.

XX AC AAG90301;

XX DT 26-SEP-2001 (first entry)

XX DE C glutamicum protein fragment SEQ ID NO: 4055.

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EPI108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 XX 18-DEC-2000; 2000EP-0127688.  
 XX  
 XX 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 XX WPI; 2001-376931/40.  
 DR N-PSDB; AAH65520.  
 XX  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX  
 PS Claim 17; SEQ ID NO: 4055; 246pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analyzing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 SQ Sequence 705 AA;  
 Query Match 35.78; Score 46; DB 22; Length 705;  
 Best Local Similarity 37.58; Pred. No. 2.3e+02;  
 Matches 9; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
 Qy 4 LEALADQTDALQSEEAAYVKADNA 27  
 Db 514 IEPYAPEADELEGESEAYKPFENA 537  
 RESULT 16  
 ID AAY54580  
 XX AAY54580 standard; Protein; 502 AA.  
 AC AAY54580;  
 XX  
 XX 25-APR-2000 (first entry)  
 DT  
 XX A rat calcium/calmodulin dependent protein kinase designated CaMK-IV.  
 DE  
 XX Calcium/calmodulin dependent protein kinase; CaMK-VI; CaMK-IV;  
 KW CaMK-VI related peptide; CARP; neuronal plasticity; epilepsy;  
 KW kainate-induced seizure; seizure; central nervous system; hippocampus;  
 KW acute brain insult; stroke; trauma; hypoxia; ischemia;  
 KW neurodegenerative disease; Alzheimer's disease.  
 XX  
 OS Rattus sp.  
 XX

FH Key Location/Qualifiers  
 FT Domain 62..324  
 FT /note= "putative catalytic domain"  
 FT Binding-site 99  
 FT /note= "residue involved in ATP binding"  
 FT Misc-difference 244  
 FT /note= "residue thought to be crucial for  
 FT activation of the kinase by phosphorylation"  
 FT Domain 227..234  
 FT /note= "Ser/Thr kinase motif"  
 FT Domain 346..364  
 FT /note= "regulatory domain"  
 XX EP978562-A1.  
 XX 09-FEB-2000.  
 PD  
 XX 06-AUG-1998; 98EP-0202654.  
 XX  
 XX 06-AUG-1998; 98EP-0202654.  
 XX (MED1-) DIV MEDICAL PHARMACOLOGY LEIDEN AMSTERDA.  
 PA  
 XX WPI; 2000-138770/13.  
 DR  
 XX Mammalian calcium/calmodulin dependent protein kinase related protein  
 XX useful as a modulator of calcium/calmodulin dependent protein kinase  
 PT activity, useful for treating seizures and damage to the central  
 PT nervous system -  
 PT  
 XX Disclosure; Page 34-36; 48pp; English.  
 XX  
 CC The present sequence represents a calcium/calmodulin dependent protein  
 CC kinase (CaMK-IV). The specification also describes another CaMK,  
 CC designated CaMK-VI, and a CaMK-VI related peptide, designated CARP. The  
 CC CARP peptide is produced by alternative splicing of the CaMK-VI gene,  
 CC and plays a role in kinase-mediated neuronal plasticity which is  
 CC associated with epilepsy. A large part of CARP (48 out of 55 amino  
 CC acids) is highly homologous with the carboxy-terminus of human  
 CC doublecortin. It is possible the calcium overload in a cell, triggered  
 CC by kainate-induced seizures, leads to activation of CaMK-VI. This  
 CC enhanced kinase activity induces CARP expression, which subsequently  
 CC acts as a modulator of the kinase, e.g. by acting as a substrate or  
 CC by irreversible binding to the activated kinase. The mammalian CARP  
 CC protein may be used for the treatment and/or prevention of seizures  
 CC (e.g. in epilepsy) and/or damage to the central nervous system,  
 CC particularly the hippocampus. CARP peptides may also be used to treat  
 CC acute brain insults (e.g. resulting from stroke, trauma, hypoxia, and  
 CC ischemia), and neurodegenerative diseases (e.g. Alzheimer's disease).  
 XX  
 SQ Sequence 502 AA;  
 Query Match 34.98; Score 45; DB 21; Length 502;  
 Best Local Similarity 58.88; Pred. No. 2.1e+02;  
 Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 DALEALADQTDALQSEEE 18  
 Db 406 EAAKAAADETMKLOSEE 422  
 RESULT 17  
 ID AAW99481  
 XX AAW99481 standard; Protein; 1362 AA.  
 AC AAW99481;  
 XX  
 XX 08-JUN-1999 (first entry)  
 DT  
 XX Murine p/CIP protein.  
 DE  
 XX Mouse; p/CIP; p300/CBP/co-integrator-associated protein; gene expression;  
 KW breast cancer; inflammatory disease; atherosclerosis; osteoporosis; ds.





PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139493.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.

PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160880.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 34.18; Score 44; DB 21; Length 274;

Best Local Similarity 39.18; Pred. No. 1.5e+02;

Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 5 EALADQTDALQSEAAVVKADNA 27

||||: ||: |:: |||

Db 53 EALQQRKALSERDKALIERDNA 75

#### RESULT 22

AAG61028

ID AAG61028 standard; Protein; 274 AA.

XX AAG61028;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 79109.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.



```
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 28-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 34.1%; Score 44; DB 21; Length 274;
Best Local Similarity 39.1%; Pred. No. 1.5e+02;
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALOSEEAAYVKADNA 27
   ||| : || : ||| : |||
Db 53 EALQQRDKALSERDKALIERDNA 75

RESULT 23
AAM78512
ID AAM78512 standard; Protein; 284 AA.
XX
AC AAM78512;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1174.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
WO200157190-A2.
PN
PD
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
WPI: 2001-476283/51.
DR N-PSDB; AAK51645.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 3417-3418; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
```

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

XX SQ Sequence 284 AA;

Query Match 34.18; Score 44; DB 22; Length 284;  
Best Local Similarity 47.68; Pred. No. 1.6e+02;  
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Oy 5 BALADQTDALQSEAEAAVVKAD 25  
||| | | | | | | | | |  
Db 62 BALKDAEDSLLAEEAAKAE 82

#### RESULT 24

AAG58292  
ID AAG58292 standard; Protein: 296 AA.

XX AC AAG58292;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 75230.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EPI033405-A2.

XX FD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126284.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 30-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 20-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147193.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157863.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 34.1%; Score 44; DB 21; Length 296;  
Best Local Similarity 39.1%; Pred. No. 1.6e+02;  
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
  
QY 5 EALADOTDALQSEEAAYVKADNA 27  
||| : || : |||  
DB 75 EALQQRDKALSRDKALIERDNA 97  
  
RESULT 25  
AAG61027  
ID AAG61027 standard; Protein; 296 AA.  
XX  
AC AAG61027;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 79108.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149436.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

34.1%; Score 44; DB 21; Length 296;

Best Local Similarity 39.1%; Pred. No. 1.6e+02;  
Matches 9; Conservative 5; Mismatches 9

QY 5 EALADQTDALQSEEA VVKADNA 27  
||| : || : |:: |||

Db 75 EALOORDKALSERDKALIERDNA 97

RESULT 26  
AAW61234  
ID AAW61234 standard; Protein: 344 AA.

XX  
XX AAW61234;  
XX  
XX  
XX  
XX DT 02-OCT-1998 (first entry)  
XX  
XX  
XX DE Streptococcus pneumoniae Spt09 protein.

Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
detection; pneumonia; otitis media; meningitis.

Streptococcus pneumoniae.

AX  
PN  
WO9818930-A2.

07-MAY-1998.

XX 30-OCT-1997; 97WO-US19422

XX  
PR 31-OCT-1996: 96US-0029960.XX  
PA (HUMA-) HUMAN GENOME SCI INC.

XX  
PI  
Choi GH, Hromockyj A, Johnson LS, Kunsch CA:

XX  
DR WPT: 1998-272224/24

DR N-PSDB; AAV27419.

xx
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis

PS Claim 11; Page 87; 118pp; English;

The present sequence represents a protein from *Streptococcus pneumoniae*. The nucleic acid sequence encoding the *Streptococcus pneumoniae* protein can be useful in vaccines for inducing protective antibodies against *Streptococcus pneumoniae*, for treatment or prevention of infection e.g. pneumonia, otitis media or meningitis. Probes based on the nucleic acid are used to detect *Streptococcus* infection (by usual hybridisation or amplification methods), also for isolating *Streptococcus* genes or their allelic variants. The protein can be used similarly to detect specific antibodies in standard immunoassays, especially for diagnosing or monitoring infections. Antibodies which bind the protein are used to detect corresponding antigens, to purify the protein and for passive immunisation (optionally coupled to a toxin). Vaccines are administered, e.g. by injection, orally or through the skin, typically at 0.01-1000 (especially 10-300)  $\mu$ mu g/ml per dose.

XY	Sequence	344 AA:
SO		

Query Match	34.1%	Score 44;	DB 19;	Length 344;
Best Local Similarity	55.0%;	Pred. NO. 1.9e+02;		
Matches 11;	Conservative	2;	Mismatches 7;	Indels 0;
				Gaps 0;

QY 8 ADQTDALQSEEA>VVKADNA 27  
| | | | | : | | | | |  
Db 3 AGOTDASOLEKA>V>SOGKA 22

RESULT 27  
AAG58291  
ID AAG58291 standard; Protein: 358 AA.

PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140931.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 34.1%; Score 44; DB 21; Length 358;  
Best Local Similarity 39.1%; Pred. No. 2e+02;  
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAAYVADNA 27  
||| : || : ||: |||  
Db 137 EALQQRDKALSRDKALIERDNA 159

RESULT 28  
AAG61026  
ID AAG61026 standard; Protein; 358 AA.  
XX  
AC AAG61026;  
XX  
DT 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 79107.  
XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 07-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 19-MAY-1999; 99US-0134768.

PR 20-MAY-1999; 99US-0134941.

PR 21-MAY-1999; 99US-0135124.

PR 24-MAY-1999; 99US-0135353.

PR 25-MAY-1999; 99US-0135629.

PR 27-MAY-1999; 99US-0136021.

PR 28-MAY-1999; 99US-0136392.

PR 01-JUN-1999; 99US-0136782.

PR 03-JUN-1999; 99US-0137222.

PR 07-JUN-1999; 99US-0137528.

PR 07-JUN-1999; 99US-0137502.

PR 08-JUN-1999; 99US-0137724.

PR 10-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 14-JUN-1999; 99US-0138847.

PR 16-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 17-JUN-1999; 99US-0139453.

PR 18-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 21-JUN-1999; 99US-0139763.

PR 22-JUN-1999; 99US-0139817.

PR 23-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 21-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 05-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 06-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 11-AUG-1999; 99US-0148171.

PR 12-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.

PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.

PR 27-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.





DR WPI: 2001-656860/75.  
XX N-PSDB; ABL15920.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure; SEQ ID NO 42243; 21pp + Sequence Listing; English.  
PS  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).  
XX  
XX SQ Sequence 421 AA;  
XX  
XX Query Match 34.1%; Score 44; DB 22; Length 421;  
XX Best Local Similarity 46.2%; Pred. No. 2.4e+02;  
XX Matches 12; Conservative 3; Mismatches 11; Indels 0; Gaps 0;  
QY 3 ALEALADQDALQSEEAAYVKADNAA 28  
ID I I I I I : I I I I I I : I I  
DB 348 AAAPADEAAPAAEAAAPPAEEAA 373  
RESULT 31  
ABB65976  
ID ABB65976 standard; Protein; 429 AA.  
XX  
AC ABB65976;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 24720.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI: 2001-656860/75.  
XX N-PSDB; ABL10079.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure; SEQ ID NO 24720; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of

CC	Insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 429 AA;
	Query Match 34.1%; Score 44; DB 22; Length 429;
	Best Local Similarity 48.1%; Pred No 2.5e+02;
	Matches 13; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
QY	2 DALEALADQTDALQSEAAVVKADNAA 28 :       :
Db	182 NAEVLAAAGNADESVAAPAGNAA 208
RESULT 32	
AAW80400	
ID	AAW80400 standard; Protein; 514 AA.
XX	
AC	AAW80400;
XX	
DT	12-JAN-1999 (first entry)
XX	
DE	A secreted protein encoded by clone dd71_2.
XX	
KW	Secreted protein; immune stimulating; suppressing;
KW	haematopoiesis regulating activity; tissue growth activity; activin;
KW	inhibin activity; chemotactic; chemokinetic activity; haemostatic;
KW	thrombolytic activity; anti-inflammatory activity; cadherin;
KW	tumour invasion suppressor activity; tumour inhibition activity.
XX	
OS	Homo sapiens.
XX	
PN	WO9840486-A2.
XX	
PD	17-SEP-1998.
XX	
Pf	13-MAR-1998; 98WO-US04977.
XX	
PR	29-OCT-1997; 97US-0960022.
PR	14-MAR-1997; 97US-0815047.
XX	
PA	(GEM ) GENETICS INST INC.
XX	
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI	Racie LA, Spaulding V, Treacy M;
XX	
DR	WPI; 1998-520812/44.
DR	N-PSDB; AAV63174.
XX	
PT	New isolated human poly:nucleotide(s) and secreted proteins -
PT	obtained from e.g. human foetal kidney, placenta, foetal brain,
PT	adult testes, adult brain or adult uterus cDNA libraries
XX	
PS	Claim 29; Pages 80-82; 110pp; English.
XX	
CC	The present sequence represents a secreted protein. The nucleic acid
CC	sequence is isolated from a human adult testes cDNA library using
CC	probe AAV63184. The polypeptide may have biological activities such as
CC	e.g. nutritional activity, immune stimulating or suppressing activity,
CC	haematopoiesis regulating activity, tissue growth activity,
CC	activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC	activity, cadherin/tumour invasion suppressor activity, tumour
CC	inhibition activity or other activities.
XX	
SQ	Sequence 514 AA;
	Query Match 34.1%; Score 44; DB 19; Length 514;

Best Local Similarity 36.0%; Pred. No. 3.1e+02;  
Matches 9; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

OY 3 ALEALADQTDALQSEEAAYVVKADNA 27  
DB 240 SIAQRQAQTOLLKEEARLVESNNA 264  
:::| | | :||| :|::|||

## RESULT 33

AAG92957  
ID AAG92957 standard; Protein: 592 AA.

XX AC AAG92957;  
XX DT 26-SEP-2001 (first entry)

XX DE C glutamicum protein fragment SEQ ID NO: 6711.  
XX KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
XX KW organic acid synthesis.

XX OS Corynebacterium glutamicum.  
XX PN EP1108790-A2.

XX PD 20-JUN-2001.  
XX PF 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.  
XX PR 07-APR-2000; 2000JP-0159162.

XX PR 03-AUG-2000; 2000JP-0280988.  
XX PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX DR WPI; 2001-376931/40.  
XX DR N-PSDB; AAH68176.

XX PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
XX PT mutation point of a gene, measuring expression of a gene, analysing  
XX PT expression profile or pattern of a gene and identifying homologous gene

XX PS Claim 17; SEQ ID NO: 6711; 246pp + Sequence Listing; English.  
XX CC The present invention provides a number of nucleotide and protein  
XX CC sequences from the Corynebacterium glutamicum. These  
XX CC are useful for identifying the mutation point of a gene derived from a  
XX CC mutant of corynebacterium, measuring expression amount and  
XX CC analysing the expression profile or expression pattern of a gene derived  
XX CC from Corynebacterium, and identifying a homologue of a gene derived  
XX CC from corynebacterium. Corynebacterium bacteria are useful for producing  
XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
XX CC particularly L-lysine. The present sequence is a protein described  
XX CC in the exemplification of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from the  
XX CC European Patent Office.

XX SQ Sequence 592 AA;  
Query Match 34.1%; Score 44; DB 22; Length 592;

Best Local Similarity 45.5%; Pred. No. 3.6e+02;  
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 SDALPALADQTDALQSEEAAYV 22  
DB 442 SDAMNALTTRDRASSERLPVI 463  
|||: ||: ||: ||:

## RESULT 34

AAW81028  
ID AAW81028 standard; Protein: 951 AA.

XX AC AAW81028;  
XX DT 12-MAY-1999 (first entry)

XX DE ER interacting domain of AIB1 protein.  
XX KW AIB1; amplified in breast cancer; cancer; steroid; receptor;  
XX KW coactivator; SCR; estrogen; ER; estrogen dependent transcription;  
XX KW breast cancer; lung cancer; colon cancer; prostate cancer;  
XX KW melanoma; PAS region.

XX OS Homo sapiens.  
XX PN WO9857982-A2.

XX PD 23-DEC-1998.  
XX PF 17-JUN-1998; 98WO-US12689.

XX PR 17-JUN-1997; 97US-0049728.  
XX PA (USGO ) US GOVERNMENT.

XX PI Meltzer P, Trent JM;  
XX DR WPI; 1999-080946/07.

XX PT New isolated steroid receptor co-activator, AIB1 - used to develop  
XX PT products for the diagnosis and treatment of steroid-responsive  
XX PT tumours, e.g. breast, lung, prostate or colon cancers or melanomas

XX PS Claim 6; Page 34-35; 57pp; English.  
XX CC The AIB1 protein is a member of the steroid receptor coactivator-1  
XX CC (SRC-1) family of nuclear receptor co-activators that interact with  
XX CC estrogen receptors (ER) to enhance ER-dependent transcription.

XX CC The AIB1 gene is amplified and over-expressed in certain cancers  
XX CC in particular breast cancer and steroid hormone responsive cancers.  
XX CC The AIB1 polypeptide can be used to identify compounds which inhibit  
XX CC ER-dependent transcription. Increased expression of the AIB1 gene  
XX CC indicates aberrantly proliferating cells, thus detection of  
XX CC increased expression of the AIB1 gene or an increase in the number  
XX CC of copies of the AIB1 gene can be used to diagnose cancer or a  
XX CC predisposition towards developing cancer. Compounds which inhibit  
XX CC expression of AIB1 or compounds which inhibit interaction of AIB1  
XX CC with steroid receptors or nuclear co-factors can be used for  
XX CC reducing the proliferation of cancer cells. The PAS domain of the  
XX CC AIB protein is a highly conserved domain among the SRC-1 family of  
XX CC proteins. The ER-interacting domain of AIB1 mediates binding of  
XX CC AIB1 with a steroid hormone receptor protein.

XX SQ Sequence 951 AA;  
Query Match 34.1%; Score 44; DB 20; Length 951;

Best Local Similarity 43.5%; Pred. No. 6.3e+02;  
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

OY 5 EALADQTDALQSEEAAYVVKADNA 27  
DB 787 QALEPKQDAFQGEAAVMMQKA 809  
:|: | | :| | :| | :

## RESULT 35

AAB59278  
ID AAB59278 standard; protein: 1412 AA.

XX AC AAB59278;  
XX DT 27-MAR-2001 (first entry)



FT /note= "encoded by TAA"  
FT Misc-difference 1474  
FT /note= "encoded by TGA"  
FT Misc-difference 1510  
FT /note= "encoded by TAG"  
FT Misc-difference 1521  
FT /note= "encoded by TGA"  
XX  
XX WO9932621-A2.  
XX  
XX 01-JUL-1999.  
XX  
XX 01-DEC-1998; 98WO-US25478.  
XX  
XX 22-DEC-1997; 97US-0068511.  
XX  
XX (AMHP ) AMERICAN HOME PROD CORP.  
XX  
XX Frall DE, Lyttle CR, Suen C;  
XX  
XX WPI; 1999-405171/34.  
XX  
XX N-PSDB; AAX80992.  
XX  
XX Nucleic acid encoding a steroid receptor co-activator-3, useful for  
XX determining the neoplastic states of cells in humans or animals  
XX  
XX Claim 2; Page 61-68; 75pp; English.  
XX  
XX This represents a human steroid receptor coactivator-3 (SRC-3). Host  
XX cells transformed with vectors comprising the SRC-3 gene can be used for  
XX the recombinant production of the SRC-3 protein. SRC-3 can be used to  
XX enhance the transcription of genes, which are regulated by hormone  
XX response elements, e.g. estrogen receptor alpha. SRC-3 can be used in  
XX methods to screen for compounds that bind to SRC-3 or which modulate  
XX formation of SRC-3 transcription complexes. Determining the amount of  
XX SRC-3 in a sample is also useful in a method to detect overexpression of  
XX SRC-3 and therefore determining the neoplastic states of cells in humans  
XX or animals.  
XX  
XX SQ Sequence 1522 AA;  
XX  
XX Query Match 34.1%; Score 44; DB 20; Length 1522;  
XX Best Local Similarity 43.5%; Pred. No. 1.1e+03;  
XX Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;  
XX  
XX Qy 5 EALADQTDALQSEAAVVKADNA 27  
XX :||:||||:||||:|  
XX Db 1086 QALEPKQAFQCGEAAVVMDOKA 1108  
XX  
XX RESULT 38  
XX ABB59874  
XX ID ABB59874 standard; Protein; 1630 AA.  
XX  
XX AC ABB59874;  
XX  
XX DT 26-MAR-2002 (first entry)  
XX  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 6414.  
XX  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX  
XX OS Drosophila melanogaster.  
XX  
XX PN WO200171042-A2.  
XX  
XX PD 27-SEP-2001.  
XX  
XX PF 23-MAR-2001; 2001WO-US09231.  
XX  
XX PR 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.  
XX  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
XX N-PSDB; ABL03977.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX  
XX Disclosure; SEQ ID NO 6414; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
XX (ABB57737-ABB72072).  
XX  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX SQ Sequence 1630 AA;  
XX  
XX Query Match 34.1%; Score 44; DB 22; Length 1630;  
XX Best Local Similarity 40.0%; Pred. No. 1.2e+03;  
XX Matches 10; Conservative 4; Mismatches 11; Indels 0; Gaps 0;  
XX  
XX Qy 4 LEALADQTDALQSEAAVVKADNAA 28  
XX :||:||||:||||:|  
XX Db 450 LNELENIODELEKEGIVIVRIDNAA 474  
XX  
XX RESULT 39  
XX AAY95962  
XX ID AAY95962 standard; Peptide; 33 AA.  
XX  
XX AC AAY95962;  
XX  
XX DT 20-NOV-2000 (first entry)  
XX  
XX DE Yeast GCN4 mutated coiled-coil peptide used in caspase assay.  
XX  
XX KW Coiled-coil; yeast; GCN4; general control of nitrogen; protease;  
XX caspase; assay; signal modulation.  
XX  
XX OS Saccharomyces cerevisiae.  
XX  
XX OS Synthetic.  
XX  
XX FH Key Location/Qualifiers  
XX FT Cleavage-site 17..18  
XX FT /note= "caspase protease cleavage site"  
XX  
XX PN WO200050635-A1.  
XX  
XX PD 31-AUG-2000.  
XX  
XX PF 18-FEB-2000; 2000WO-GB00601.  
XX  
XX PR 25-FEB-1999; 99GB-0004401.  
XX  
XX (FLUO-) FLUORESCENCE LTD.  
XX  
XX PI Coyler J, Lightowler J;  
XX  
XX WPI; 2000-565459/52.  
XX  
XX Labeled polypeptide binding partner compositions useful for monitoring  
XX protease activity by detecting signal modulation -

```

XX Example 5; Page 59; 90pp; English.
XX
CC The present sequence is that of a homomeric coiled-coil domain
CC derived from amino acids 249-281 of Saccharomyces cerevisiae GCN4
CC (general control of nitrogen). The native GCN4 sequence was
CC modified to include a caspase site. A polypeptide useful in the
CC present invention is capable of multimerising with similar or
CC different polypeptides. The polypeptide requires a binding domain,
CC such as a coiled-coil domain, which allows multimerisation. The
CC multimer preferably comprises a first polypeptide associated with a
CC label and a second polypeptide, where (a) at least 1 of the
CC polypeptides is susceptible to protease digestion, (b) association
CC of the polypeptides to form a multimer is detectable via a signal
CC emitted by the signal, and (c) digestion of at least 1 polypeptides
CC results in dissociation of the multimer and modulation of the signal
CC emitted by the label. The multimer is used to monitor the activity
CC of a protease. In an example of the method, the polypeptide
CC incorporates the present GCN4 coiled-coil domain and is used in a
CC caspase assay, with the signal analysed by fluorescence resonance
CC energy transfer.
XX
SQ Sequence 33 AA;
Query Match 33.3%; Score 43; DB 21; Length 33;
Best Local Similarity 40.9%; Pred. No. 18;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
QY 2 DALEALADQTDALQSEEAAYVK 23
DB 7 DKVEELLDNDHLENEVACLKK 28
RESULT 40
AAY44712
ID AAY44712 standard; Protein; 92 AA.
XX
AC AAY44712;
XX
DT 25-APR-2000 (first entry)
XX
DE Shorthorn sculpin skin-type antifreeze protein.
XX
KW clone s3-2; shorthorn sculpin skin-type antifreeze protein; sssAFP;
KW antibacterial; cold tolerance; ice recrystallisation;
KW refrigerated food; antifreeze.
XX
OS Myoxocephalus scorpius.
XX
PN W0200000512-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-CA00601.
XX
PR 26-JUN-1998; 98US-0090794.
PR 07-AUG-1998; 98US-0095713.
PR 24-JUN-1999; 99US-0344529.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Hew CL;
XX
DR WPI: 2000-170905/15.
DR N-PSDB; AA249960.
XX
XX Novel antifreeze polypeptides and polynucleotides used to make cells
PT cold resistant and to improve the palatability of cold foods and
PT liquids.
XX
XX Example 1; Fig 1; 61pp; English.
XX
XX The present sequence is shorthorn sculpin skin-type antifreeze protein

```

```

CC (sssAFP-2) encoded by cDNA clone s3-2. The AFPs are alanine-rich
CC polypeptides that are synthesised in the peripheral tissues, such as
CC skin and gills, and serve to depress the freezing temperature of
CC intracellular fluids. The AFPs are used to make aqueous compositions
CC resistant to freezing, to improve the cold tolerance, for inhibiting ice
CC recrystallisation, and for stabilising biological membranes. They are
CC also used to improve the shelf life and food quality of many
CC refrigerated foods and have antibacterial properties. This has many
CC useful applications in medicine, agriculture and aquaculture.
XX
SQ Sequence 92 AA;
Query Match 33.3%; Score 43; DB 21; Length 92;
Best Local Similarity 39.3%; Pred. No. 59;
Matches 11; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
QY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
DB 32 SAAAAAIAIAEAAEAEEAAATKSNVA 59

```

Search completed: October 13, 2002, 02:08:16  
Job time : 13.3797 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 02:06:51 ; Search time 24.6578 Seconds  
(without alignments)  
157.503 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSIEQPRIIPGGTILNL.....KKEERQNVNLTPTGSESNP 159

Scoring table: BLOSUM62

Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PT05\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	91	11.7	413	US-08-669-408B-10	Sequence 10, Appl
2	85	10.9	664	US-08-669-408B-2	Sequence 2, Appl
3	76	9.8	50	US-08-294-189-14	Sequence 14, Appl
4	76	9.8	3248	US-08-353-700-1	Sequence 1, Appl
5	76	9.8	3248	PCT-US95-16216-1	Sequence 1, Appl
6	75	9.7	2482	US-08-328-254-6	Sequence 6, Appl
7	69	8.9	46	US-08-363-311-25	Sequence 25, Appl
8	69	8.9	46	US-08-463-288A-25	Sequence 25, Appl
9	69	8.9	46	US-08-470-445A-25	Sequence 25, Appl
10	69	8.9	46	US-08-462-679-25	Sequence 25, Appl
11	69	8.9	46	US-08-466-210A-25	Sequence 25, Appl
12	69	8.9	46	US-08-467-147A-25	Sequence 25, Appl
13	69	8.9	46	US-08-469-014-25	Sequence 25, Appl
14	69	8.9	46	US-09-346-290-25	Sequence 25, Appl
15	69	8.9	46	PCT-US93-10506A-25	Sequence 25, Appl
16	69	8.9	46	PCT-US93-10506-25	Sequence 25, Appl
17	68	8.8	1786	US-08-973-463-8	Sequence 8, Appl
18	67	8.6	472	US-08-216-894-10	Sequence 10, Appl
19	67	8.6	472	US-09-115-746-10	Sequence 10, Appl
20	66	8.5	46	US-08-363-311-26	Sequence 26, Appl
21	66	8.5	46	US-08-463-288A-26	Sequence 26, Appl
22	66	8.5	46	US-08-470-445A-26	Sequence 26, Appl
23	66	8.5	46	US-08-462-679-26	Sequence 26, Appl
24	66	8.5	46	US-08-466-210A-26	Sequence 26, Appl
25	66	8.5	46	US-08-467-147A-26	Sequence 26, Appl
26	66	8.5	46	US-08-469-014-26	Sequence 26, Appl
27	66	8.5	46	US-09-346-290-26	Sequence 26, Appl

28 8.5 46 5 PCT-US93-10506A-26 Sequence 26, Appl  
29 8.5 46 5 PCT-US93-10506-26 Sequence 26, Appl  
30 8.5 593 2 US-08-591-079-8 Sequence 8, Appl  
31 8.5 593 2 US-08-591-079-10 Sequence 10, Appl  
32 8.4 1162 2 US-08-728-323A-2 Sequence 2, Appl  
33 64 8.2 564 2 US-08-216-894-2 Sequence 2, Appl  
34 64 8.2 564 4 US-09-115-746-2 Sequence 2, Appl  
35 64 8.2 643 2 US-08-216-894-8 Sequence 8, Appl  
36 64 8.2 643 4 US-09-115-746-8 Sequence 8, Appl  
37 64 8.2 1111 1 US-08-317-450B-15 Sequence 15, Appl  
38 64 8.2 1111 4 US-08-800-593-15 Sequence 15, Appl  
39 64 8.2 1193 1 US-08-317-450B-13 Sequence 13, Appl  
40 64 8.2 1193 4 US-08-800-593-13 Sequence 13, Appl  
41 63 8.1 288 2 US-08-961-858-6 Sequence 6, Appl  
42 63 8.1 288 3 US-09-089-593-6 Sequence 6, Appl  
43 63 8.1 288 3 US-08-950-925-4 Sequence 4, Appl  
44 62 8.0 461 4 US-09-346-408-8 Sequence 8, Appl  
45 62 8.0 999 2 US-08-770-301A-3 Sequence 3, Appl

#### ALIGNMENTS

#### RESULT 1

US-08-669-408B-10  
; Sequence 10, Application US/08669408B  
; Patent No. 6100055  
; GENERAL INFORMATION:  
; APPLICANT: GUSS, Bengt  
; APPLICANT: JONSSON, Hans  
; APPLICANT: LINDBERG, Martin  
; APPLICANT: MUELLER, Hans-Peter  
; APPLICANT: RANTAMAKI, Liisa K.  
; TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING  
; TITLE OF INVENTION: PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/669,408B  
; FILING DATE: 03-JUL-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/SE94/00826  
; FILING DATE: 06-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: SE 9302855-3  
; FILING DATE: 06-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 61743/102  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 413 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-669-408B-10

Query Match	11.7%	Score 91	DB 3	Length 413
Best Local Similarity	59.4%	Pred. No.	0.0021	
Matches 19	Conservative	1	Mismatches 12	Indels 0
			Gaps	0

**Qy** 128 DVKKTGDKTPVWKKEERQNVTLPITGEESNP 159  
+ + + + + : + + + + +  
**Db** 357 DTKKEDAKKPEAKEEAKKAATLPTTGEESNP 388

```

: RESULT 2
: US-08-669-408B-2
: ; Sequence 2, Application US/08669408B
: ; Patent No. 6100055
: ; GENERAL INFORMATION:
: ; APPLICANT: GUSS, Bengt
: ; APPLICANT: JONSSON, Hans
: ; APPLICANT: LINDBERG, Martin
: ; APPLICANT: MUELLER, Hans-Peter
: ; APPLICANT: RANTAMAKI, Liisa K.
: ; TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING
: ; TITLE OF INVENTION: PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS
: ; NUMBER OF SEQUENCES: 12
: ; CORRESPONDENCE ADDRESS:
:

```

Query Match	10.9%	Score 85	DB 3	Length 664
Best Local Similarity	56.2%	Pred. No.	0.021	
Matches 18	Conservative	2	Mismatches	12
			Indels	0
			Gaps	0

QY 128 DVKKT KDTPVVKKEERQNVNTLP TTGEESNP 159  
| | | : | | | | : | | | | | | | |  
Db 608 DAKKPEAKKEEAKKEAKKAATLP TTGEESNP 639

RESULT 3  
US-08-294-189-14  
; Sequence 14, Application US/08294189

```

/ Patent No. 5707822
/ GENERAL INFORMATION:
/ APPLICANT: Fischetti, Vincent A.
/ APPLICANT: Rakonjac, Jasna
/ APPLICANT: Robbins, John
/ TITLE OF INVENTION: GENE SERUM OPACITY FACTOR
/ NUMBER OF SEQUENCES: 23
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Klauber & Jackson
/ CITY: Hackensack
/ STATE: New Jersey
/ COUNTRY: USA
/ ZIP: 07601
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/294,189
/ FILING DATE: 22-AUG-1994
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Jackson Esq., David A.
/ REGISTRATION NUMBER: 26,742
/ REFERENCE/DOCKET NUMBER: 600-1-118
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201 487-5800
/ TELEFAX: 201 343-1684
/ TELEX: 133521
/ INFORMATION FOR SEQ ID NO: 14:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 50 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: unknown
/ MOLECULE TYPE: peptide
/ HYPOTHEICAL: NO
/ FRAGMENT TYPE: C-terminal
/ ORIGINAL SOURCE:
/ ORGANISM: Group G Streptococci
/ US-08-294-189-14

Query Match 9.8%; Score 76; DB 1; Length 50;
Best Local Similarity 58.3%; Pred. No. 0.011;
Matches 14; Conservative 3; Mismatches 7; Indels 0;

QY 136 KPVVKKEERONVNTLPTTGESNP 159
|| ||: : ||||| |||
DB 2 KPEAKDDAKKAETLPTTGESNP 25

RESULT 4
US-08-353-700-1
/ Sequence 1, Application US/08353700
/ Patent No. 5599919
/ GENERAL INFORMATION:
/ APPLICANT: YEN, TIMOTHY J.
/ APPLICANT: RATTNER, JEROME B.
/ TITLE OF INVENTION: NUCLEIC ACID ENCODING A
/ TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
/ TITLE OF INVENTION: AND METHODS OF USE
/ NUMBER OF SEQUENCES: 4
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
/ STREET: 1601 MARKET STREET, SUITE 720
/ CITY: PHILADELPHIA
/ STATE: PA
/ COUNTRY: USA
/ ZIP: 19103-2307
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk

```

```

RESULT 4
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 559919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```



COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/353,700  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: REED, JANET E.  
REGISTRATION NUMBER: 36,252  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3248 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: HUMAN  
US-08-353-700-1

Query Match 9.8%; Score 76; DB 1; Length 3248;  
Best Local Similarity 21.0%; Pred. No. 2.2;  
Matches 21; Conservative 28; Mismatches 51; Indels 0; Gaps 0;  
QY 46 EDKEATTATEAASSDALEALADOTDALQSEAAVVKADNAASDALEALADOTDALQSEEA 105  
DB 2520 EXEQEKVQMKESSTAMEIQTOLKELNERYAALHNDQEKACKKEONLSQVECLELEKA 2579  
QY 106 EYVQSDNAASDAWEKAATPIALDVKKTKDPKPVVKKEERQ 145  
DB 2580 QLLQGLDEAKNNIVLVQSSVKGLIQEVEDGKQKLEKDEE 2619

## RESULT 5

PCT-US95-16216-1  
Sequence 1, Application PC/TUS9516216  
GENERAL INFORMATION:  
APPLICANT: Yen, Timothy J.  
APPLICANT: Ratiner, Jerome B.  
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
Expressed Kinetochores Protein, and Methods of Use  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
STREET: 1601 Market Street Suite 720  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/16216  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/353,700  
FILING DATE: 09-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Reed, Janet E.  
REGISTRATION NUMBER: 36,252  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3248 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PCT-US95-16216-1

Query Match 9.8%; Score 76; DB 5; Length 3248;  
Best Local Similarity 21.0%; Pred. No. 2.2;  
Matches 21; Conservative 28; Mismatches 51; Indels 0; Gaps 0;  
QY 46 EDKEATTATEAASSDALEALADOTDALQSEAAVVKADNAASDALEALADOTDALQSEEA 105  
DB 2520 EXEQEKVQMKESSTAMEIQTOLKELNERYAALHNDQEKACKKEONLSQVECLELEKA 2579  
QY 106 EYVQSDNAASDAWEKAATPIALDVKKTKDPKPVVKKEERQ 145  
DB 2580 QLLQGLDEAKNNIVLVQSSVKGLIQEVEDGKQKLEKDEE 2619

## RESULT 6

US-08-328-254-6  
Sequence 6, Application US/08328254  
Patent No. 5710022  
GENERAL INFORMATION:  
APPLICANT: Zhu, Xueliang  
APPLICANT: Lee, Wen-Hwa  
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/328,254  
FILING DATE: 24-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/141,239  
FILING DATE: 22-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-CJ 1191  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2482 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-328-254-6

Query Match 9.7%; Score 75; DB 1; Length 2482;  
Best Local Similarity 21.0%; Pred. No. 2.1;  
Matches 21; Conservative 28; Mismatches 51; Indels 0; Gaps 0;

QY 46 EDKEATTATEAASSDALEALADOTDALQSEAAVVKADNAASDALEALADOTDALQSEEA 105  
DB 2520 EXEQEKVQMKESSTAMEIQTOLKELNERYAALHNDQEKACKKEONLSQVECLELEKA 2579  
QY 106 EYVQSDNAASDAWEKAATPIALDVKKTKDPKPVVKKEERQ 145  
DB 2580 QLLQGLDEAKNNIVLVQSSVKGLIQEVEDGKQKLEKDEE 2619



```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,445A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363,311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.237000A
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: peptide
; US-08-470-445A-25

Query Match 8.9%; Score 69; DB 2; Length 46;
Best Local Similarity 65.0%; Pred. No. 0.075;
Matches 13; Conservative 2; Mismatches 5; Indels 5; Gaps 0;

OY 140 KKEERQNVNLTPTTGESNP 159
   |||: : ||||| |||
Db 2 KKEDAKKAETLPTTGESNP 21

RESULT 10
US-08-462-679-25
; Sequence 25, Application US/08462679
; Patent No. 5847081
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; TITLE OF INVENTION: Conjugate Vaccine For Group B
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,679
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363,311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.2370008
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: peptide
; US-08-462-679-25

Query Match 8.9%; Score 69; DB 2; Length 46;
Best Local Similarity 65.0%; Pred. No. 0.075;
Matches 13; Conservative 2; Mismatches 5; Indels 5; Gaps 0;

OY 140 KKEERQNVNLTPTTGESNP 159
   |||: : ||||| |||
Db 2 KKEDAKKAETLPTTGESNP 21

RESULT 11
US-08-466-210A-25
; Sequence 25, Application US/08466210A
; Patent No. 5858362
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine For Group B
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,210A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363,311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.237000B
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK

```

; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 46 amino acids  
; TYPE: amino acid  
; TOPOLOGY: both  
; MOLECULE TYPE: peptide  
US-08-466-210A-25

Query Match 8.9%; Score 69; DB 2; Length 46;  
Best Local Similarity 65.0%; Pred. No. 0.075;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNTLPTTGESNP 159  
DB 2 KKEDAKKAETLPTTGESNP 21  
||||: ||||||| |||

## RESULT 12

US-08-467-147A-25  
; Sequence 25, Application US/08467147A  
; Patent No. 5908629  
; GENERAL INFORMATION:  
; APPLICANT: Michel, James L.  
; APPLICANT: Kasper, Dennis L.  
; APPLICANT: Ausubel, Frederick C.  
; APPLICANT: Madoff, Lawrence C.  
; TITLE OF INVENTION: Conjugate Vaccine for Group B  
; NUMBER OF INVENTIONS: 65  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, NW, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,147A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/363,311  
; FILING DATE: 22-DEC-1994  
; PRIOR APPLICATION NUMBER: US 07/968,866  
; FILING DATE: 02-NOV-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/408,036  
; FILING DATE: 15-SEP-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bugalsky, Lawrence B.  
; REGISTRATION NUMBER: 35,086  
; REFERENCE/DOCKET NUMBER: 0609.2370009  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; TELEX: 248636 SSK  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 46 amino acids  
; TYPE: amino acid  
; TOPOLOGY: both  
; MOLECULE TYPE: peptide  
US-08-467-147A-25

Query Match 8.9%; Score 69; DB 2; Length 46;  
Best Local Similarity 65.0%; Pred. No. 0.075;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNTLPTTGESNP 159  
DB 2 KKEDAKKAETLPTTGESNP 21  
||||: ||||||| |||

## RESULT 13

US-08-469-014-25  
; Sequence 25, Application US/08469014  
; Patent No. 5968521  
; GENERAL INFORMATION:  
; APPLICANT: Michel, James L.  
; APPLICANT: Kasper, Dennis L.  
; APPLICANT: Ausubel, Frederick M.  
; APPLICANT: Madoff, Lawrence C.  
; TITLE OF INVENTION: Conjugate Vaccine Against Group B  
; NUMBER OF INVENTIONS: 65  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, NW, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,014  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/363,311  
; FILING DATE: 22-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/968,866  
; FILING DATE: 02-NOV-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/408,036  
; FILING DATE: 15-SEP-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bugalsky, Lawrence B.  
; REGISTRATION NUMBER: 35,086  
; REFERENCE/DOCKET NUMBER: 0609.2370006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; TELEX: 248636 SSK  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 46 amino acids  
; TYPE: amino acid  
; TOPOLOGY: both  
; MOLECULE TYPE: peptide  
US-08-469-014-25

Query Match 8.9%; Score 69; DB 2; Length 46;  
Best Local Similarity 65.0%; Pred. No. 0.075;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNTLPTTGESNP 159  
DB 2 KKEDAKKAETLPTTGESNP 21  
||||: ||||||| |||

## RESULT 14

US-09-346-290-25  
; Sequence 25, Application US/09346290  
; Patent No. 6342223  
; GENERAL INFORMATION:

```
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine Against Group B
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; STREET: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/346,290
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/469,014
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.2370006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2500
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: peptide
; US-09-346-290-25

Query Match      8.9%; Score 69; DB 4; Length 46;
Best Local Similarity 65.0%; Pred. No. 0.075;
Matches 13; Conservative 2; Mismatches 5; Indels 5; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159
   ||| : ||||| |||
Db 2 KKEDAKKAETLPTTGESNP 21

RESULT 15
PCT-US93-10506A-25
; Sequence 25, Application PC/TUS9310506A
; GENERAL INFORMATION:
; APPLICANT: THE GENERAL HOSPITAL CORPORATION
; APPLICANT: Fruit Street
; APPLICANT: Boston, Massachusetts 02114
; APPLICANT: United States of America
; APPLICANT: 75 Francis Street
; APPLICANT: Boston, Massachusetts 02115
; APPLICANT: United States of America
; TITLE OF INVENTION: Conjugate Vaccine Against Group B
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
```

```
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10506A
; FILING DATE: 02-NOV-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 0609.237PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: peptide
; PCT-US93-10506A-25

Query Match      8.9%; Score 69; DB 5; Length 46;
Best Local Similarity 65.0%; Pred. No. 0.075;
Matches 13; Conservative 2; Mismatches 5; Indels 5; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159
   ||| : ||||| |||
Db 2 KKEDAKKAETLPTTGESNP 21

RESULT 16
PCT-US93-10506-25
; Sequence 25, Application PC/TUS9310506
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine Against Group B
; TITLE OF INVENTION: Streptococcus
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W.; Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10506
; FILING DATE: 02-NOV-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; ATTORNEY/AGENT INFORMATION:
```

[illegible]

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 472 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-115-746-10

Query Match 8.5%; Score 67; DB 4; Length 472;  
Best Local Similarity 20.0%; Pred. No. 2.6;  
Matches 18; Conservative 20; Mismatches 52; Indels 0; Gaps 0;

QY 33 EERAIDELKKQATEDEKATTAETAASSDALEALADOTDALOSEAAVVKADNAASDALEA 92  
DB 255 EKORAAEATKVAEAEKOKAAEATKVAEAEKORAAEATKVAEAEKOKAAEATKVAGDEKQK 314  
QY 93 LADOTDALOSEAAEVVQSDNAASDANEKAA 122  
DB 315 AAEATKVAEAEKOKAAEATKVAEAEKOKAA 344

RESULT 20  
US-08-363-311-26  
Sequence 26, Application US/08363311  
Patent No. 6548241  
GENERAL INFORMATION:  
APPLICANT: Michel, James L.  
APPLICANT: Ausubel, Frederick M.  
TITLE OF INVENTION: Conjugate Vaccine Against Group B  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-2678  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/363,311  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/968,866  
FILING DATE: 02-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Cimbala, Michele A.  
REGISTRATION NUMBER: 33,851  
REFERENCE/DOCKET NUMBER: 0609.3740004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
TELEX: 248636 SSK  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-08-363-311-26

Query Match 8.5%; Score 66; DB 1; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159  
DB 2 KDDAKKAETLPTTGESNP 21  
RESULT 21  
US-08-463-288A-26  
Sequence 26, Application US/08463288A  
Patent No. 5820860  
GENERAL INFORMATION:  
APPLICANT: Michel, James L.  
APPLICANT: Kasper, Dennis L.  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Madoff, Lawrence C.  
TITLE OF INVENTION: Conjugate Vaccine For Group B  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, NW, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/463,288A  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/363,311  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/968,866  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/408,036  
FILING DATE: 15-SEP-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609.2370007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
TELEX: 248636 SSK  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-08-463-288A-26

Query Match 8.5%; Score 66; DB 2; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159  
DB 2 KDDAKKAETLPTTGESNP 21

RESULT 22  
US-08-470-445A-26  
Sequence 26, Application US/08470445A  
Patent No. 5843444  
GENERAL INFORMATION:

```

; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine Against Group B
; TITLE OF INVENTION: Streptococcus
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470.445A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363.311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968.866
; FILING DATE: 02-NOV-1992
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.237000A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 26:
; APPLICATION NUMBER: US 07/408.036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.237000A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 26:
; APPLICATION NUMBER: US 07/408.036
; FILING DATE: 15-SEP-1989
; TYPE: amino acid
; MOLECULE TYPE: both
; US-08-470-445A-26

Query Match      8.5%; Score 66; DB 2; Length 46;
Best Local Similarity 60.0%; Pred. No. 0.18;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEQRQNTLPTTGEESNP 159
Db 2 KKDDAKKAETLPTTGEESNP 21
|||: : ||||| |||

RESULT 23
US-08-462-679-26
; Sequence 26, Application US/08462679
; Patent No. 5847081
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine For Group B
; TITLE OF INVENTION: Streptococcus
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington

```

```

; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462.679
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363.311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968.866
; FILING DATE: 02-NOV-1992
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.237000B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 26:
; APPLICATION NUMBER: US 07/408.036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.237000B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 26:
; APPLICATION NUMBER: US 07/408.036
; FILING DATE: 15-SEP-1989
; TYPE: amino acid
; MOLECULE TYPE: both
; US-08-462-679-26

Query Match      8.5%; Score 66; DB 2; Length 46;
Best Local Similarity 60.0%; Pred. No. 0.18;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEQRQNTLPTTGEESNP 159
Db 2 KKDDAKKAETLPTTGEESNP 21
|||: : ||||| |||

RESULT 24
US-08-466-210A-26
; Sequence 26, Application US/08466210A
; Patent No. 5858362
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine For Group B
; TITLE OF INVENTION: Streptococcus
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466.210A
; FILING DATE: 06-JUN-1995

```



CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION NUMBER: US 08/363,311  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION NUMBER: US 07/968,866  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION NUMBER: US 07/408,036  
FILING DATE: 15-SEP-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609,237000B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-08-466-210A-26

Query Match 8.5%; Score 66; DB 2; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.16;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 140 KKEERQNVNTLPTTGESNP 159  
|||: : ||||| |||  
Db 2 KKDDAKKAETLPTTGESNP 21

## RESULT 25

US-08-467-147A-26  
Sequence 26, Application US/08467147A  
Patent No. 5908529  
GENERAL INFORMATION:  
APPLICANT: Michel, James L.  
APPLICANT: Kasper, Dennis L.  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Madoff, Lawrence C.  
TITLE OF INVENTION: Conjugate Vaccine For Group B  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, NW, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,147A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/363,311  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/968,866  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/408,036  
FILING DATE: 15-SEP-1989  
ATTORNEY/AGENT INFORMATION:

NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609,2370009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-08-467-147A-26

Query Match 8.5%; Score 66; DB 2; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.16;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 140 KKEERQNVNTLPTTGESNP 159  
|||: : ||||| |||  
Db 2 KKDDAKKAETLPTTGESNP 21

## RESULT 26

US-08-469-014-26  
Sequence 26, Application US/08469014  
Patent No. 5968521  
GENERAL INFORMATION:  
APPLICANT: Michel, James L.  
APPLICANT: Kasper, Dennis L.  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Madoff, Lawrence C.  
TITLE OF INVENTION: Conjugate Vaccine Against Group B  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, NW, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,014  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/363,311  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/968,866  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/408,036  
FILING DATE: 15-SEP-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609,2370006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid

; TOPOLOGY: both  
; MOLECULE TYPE: peptide  
; US-08-469-014-26

Query Match 8.5%; Score 66; DB 2; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159  
|:::| | | | | | | | | |  
Db 2 KKDAAKKAETLPTTGESNP 21

## RESULT 27

US-09-346-290-26  
; Sequence 26, Application US/09346290  
; Patent No. 6342223  
; GENERAL INFORMATION:  
; APPLICANT: Michel, James L.  
; APPLICANT: Kasper, Dennis L.  
; APPLICANT: Ausubel, Frederick M.  
; APPLICANT: Madoff, Lawrence C.  
; TITLE OF INVENTION: Conjugate Vaccine Against Group B  
; TITLE OF INVENTION: Conjugate Vaccine Against Group B  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, NW, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/346,290  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA: US/08/469,014  
APPLICATION NUMBER: 08/469,014  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/968,866  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/408,036  
FILING DATE: 13-SEP-1989

ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609.2370006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid

TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-09-346-290-26

Query Match 8.5%; Score 66; DB 4; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159  
|:::| | | | | | | | | |  
Db 2 KKDAAKKAETLPTTGESNP 21

## RESULT 28

PCT-US93-10506A-26  
; Sequence 26, Application PC/TUS9310506A  
; GENERAL INFORMATION:  
; APPLICANT: THE GENERAL HOSPITAL CORPORATION  
; APPLICANT: Fruit Street  
; APPLICANT: Boston, Massachusetts 02114  
; APPLICANT: United States of America  
; APPLICANT: 75 Francis Street  
; APPLICANT: Boston, Massachusetts 02115  
; APPLICANT: United States of America  
; TITLE OF INVENTION: Conjugate Vaccine Against Group B  
; TITLE OF INVENTION: Streptococcus  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Ave., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10506A  
FILING DATE: 02-NOV-1993  
CLASSIFICATION:

PRIOR APPLICATION DATA: US/07/968,866  
APPLICATION NUMBER: 07/968,866  
FILING DATE: 02-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Cimbala, Michele A.  
REGISTRATION NUMBER: 33,851  
REFERENCE/DOCKET NUMBER: 0609.237PC01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
PCT-US93-10506A-26

Query Match 8.5%; Score 66; DB 5; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159  
|:::| | | | | | | | | |  
Db 2 KKDAAKKAETLPTTGESNP 21

## RESULT 29

PCT-US93-10506-26  
; Sequence 26, Application PC/TUS9310506  
; GENERAL INFORMATION:  
; APPLICANT: Michel, James L.  
; APPLICANT: Kasper, Dennis  
; APPLICANT: Ausubel, Frederick M.  
; APPLICANT: Madoff, Lawrence C.  
; TITLE OF INVENTION: Conjugate Vaccine Against Group B  
; TITLE OF INVENTION: Streptococcus  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

```

; TELEFAX: (202) 887-0764
; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 593 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-591-079-8

Query Match 8.5%; Score 66; DB 2; Length 593;
Best Local Similarity 26.7%; Pred. No. 4.6;
Matches 23; Conservative 12; Mismatches 51; Indels 0; Gaps 0;

QY 58 SDALEALADQDTALQSEEAAYVKADNAASDALEALADQDTALQSEEAAYVQSDNAASDA 117
    ::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 139 ATDLIYASIKTTAKSVYDRAATYKLTQAQNKLOSLLDPADPGYAAQAAAYEQAGKEATEA 198
    ::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY 118 WEKAATPIALDVKKTKDTKPKVKEE 143
    |::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 199 KEALDKATDATVKGTTDAKAKAEKAD 224
    |::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 31
US-08-591-079-10
; Sequence 10, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: zychlinksky, Arturo
; APPLICANT: Chen, Yajing
; TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
; NUMBER OF SEQUENCES: 10

```

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MORRISON & FOERSTER  
 STREET: 2000 Pennsylvania Avenue, NW  
 CITY: Washington

STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:

[illegible]

Db 199 KEALDKATDATYKAGTDAKAKAKAD 224

RESULT 32

US-08-728-323A-2

Sequence 2, Application US/08728323A

Patent No. 5948676

GENERAL INFORMATION:

APPLICANT: Chang, Yuan

APPLICANT: Bohenzky, Roy A.

APPLICANT: Russo, James J.

APPLICANT: Edelman, Isidore S.

APPLICANT: Moore, Patrick S.

TITLE OF INVENTION: Immediate Early Protein From Kaposi's

TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA

TITLE OF INVENTION: Encoding Same And Uses Thereof

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/728,323A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION/DOCKET NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS

TELEPHONE: 212-278-0400

TELEFAX: 212-391-0525

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1162 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-728-323A-2

Query Match 8.4%; Score 65; DB 2; Length 1162;

Best Local Similarity 15.5%; Pred. No. 14;

Matches 18; Conservative 35; Mismatches 63; Indels 0; Gaps 0;

QY 30 LRNEERAIDELKKQATDEKATTAIEAASSDALEALADDTALQSEEAHVKNADNAA 89

Db 776 LEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQ 835

QY 90 LEALADDTALQSEEAHVQSDNAAADWEKAATPIALDVKKTKDPKPVVKKPEEQ 145

Db 836 EQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQ 891

RESULT 33

US-08-216-894-2

Sequence 2, Application US/08216894

Patent No. 5876734

GENERAL INFORMATION:

APPLICANT: Kirchhoff, Louis V.

APPLICANT: Otsu, Keiko

TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION

TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/115,746

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/216,894

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Bent, Stephen A.

REGISTRATION/DOCKET NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 85326/102/DRLO

QY 29 ALRNEERAIDELKKQATDEKATTAIEAASSDALEALADDTALQSEEAHVKNADNAA 88

Db 367 AVETEKQRAAEATKVAEAEKQKAAAEATKVAEAEKQKAA 122

QY 89 ALEALADDTALQSEEAHVQSDNAAADWEKAA 122

Db 427 EKQKAAEATKVAEAEKQKAAAEATKVAEAEKQKAA 460

RESULT 34

US-09-115-746-2

Sequence 2, Application US/09115746

Patent No. 6228601

GENERAL INFORMATION:

APPLICANT: Kirchhoff, Louis V.

APPLICANT: Otsu, Keiko

TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION

TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/115,746

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/216,894

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Bent, Stephen A.

REGISTRATION/DOCKET NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 85326/102/DRLO

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-115-746-2

Query Match      8.2%; Score 64; DB 4; Length 564;
Best Local Similarity 20.2%; Pred. No. 7.6;
Matches 19; Conservative 20; Mismatches 55; Indels 0; Gaps 0;

QY 29 ALRNEERAIDELKKQAIEDKEATTAIEAASSDSDALEALADOTDLOLSEEAIVVKADNNAASD 88
DB 367 AVTEKQRAAEATKVAEAEKQKAAEAAKAVETEKQRAAEATKVAEAEKQRAAEAMKVAAE 426

QY 89 ALEALADOTDLOLSEEAIVVQSDNNAASDAWEKAA 122
DB 427 EKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAA 460

RESULT 35
US-08-216-894-8
; Sequence 8, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchoff, Louis V.
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-894-8

Query Match      8.2%; Score 64; DB 2; Length 643;
Best Local Similarity 20.2%; Pred. No. 9;
Matches 19; Conservative 20; Mismatches 55; Indels 0; Gaps 0;

QY 29 ALRNEERAIDELKKQAIEDKEATTAIEAASSDSDALEALADOTDLOLSEEAIVVKADNNAASD 88
DB 367 AVTEKQRAAEATKVAEAEKQKAAEAAKAVETEKQRAAEATKVAEAEKQRAAEAMKVAAE 426
```

```
QY 89 ALEALADOTDLOLSEEAIVVQSDNNAASDAWEKAA 122
DB 427 EKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAA 460

RESULT 36
US-09-115-746-8
; Sequence 8, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchoff, Louis V.
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-115-746-8

Query Match      8.2%; Score 64; DB 4; Length 643;
Best Local Similarity 20.2%; Pred. No. 9;
Matches 19; Conservative 20; Mismatches 55; Indels 0; Gaps 0;

QY 29 ALRNEERAIDELKKQAIEDKEATTAIEAASSDSDALEALADOTDLOLSEEAIVVKADNNAASD 88
DB 367 AVTEKQRAAEATKVAEAEKQKAAEAAKAVETEKQRAAEATKVAEAEKQRAAEAMKVAAE 426

QY 89 ALEALADOTDLOLSEEAIVVQSDNNAASDAWEKAA 122
DB 427 EKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAA 460

RESULT 37
US-08-317-450B-15
; Sequence 15, Application US/08317450B
; Patent No. 5660982
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; TITLE OF INVENTION: Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: BANNER & ALLEGRETTI, LTD.  
STREET: Ten South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317.450B  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Chao, Mark  
REGISTRATION NUMBER: 37,293  
REFERENCE/DOCKET NUMBER: 94,778  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1111 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-450B-15

Query Match 8.2%; Score 64; DB 1; Length 1111;  
Best Local Similarity 32.3%; Pred. No. 18;  
Matches 20; Conservative 8; Mismatches 34; Indels 0; Gaps 0;

QY 30 LNEERAIDELKKQATIEDKEATTATEAASSDALEALADOTDALQSEEAHVVKADNAASDA 89  
DB 955 LREFDLOVNRKAEAEAMKRLSYISQKVSADSKTQQAERALGSAADAQRAKNKGAE 1014  
QY 90 LE 91  
DB 1015 LE 1016

RESULT 38  
US-08-800-593-15  
Sequence 15, Application US/08800593  
Patent No. 6143505  
GENERAL INFORMATION:  
APPLICANT: Tryggvason, Karl  
APPLICANT: Kallunki, Pekka  
APPLICANT: Pyke, Charles  
TITLE OF INVENTION: Laminin Chains: Diagnostic and  
TITLE OF INVENTION: Therapeutic Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
STREET: 300 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800.593  
FILING DATE: 18-FEB-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/317.450  
FILING DATE: 04-OCT-1994

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Chao, Mark  
REGISTRATION NUMBER: 37,293  
REFERENCE/DOCKET NUMBER: 94,778-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-913-0001  
TELEFAX: 312-913-0002  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1111 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-800-593-15

Query Match 8.2%; Score 64; DB 4; Length 1111;  
Best Local Similarity 32.3%; Pred. No. 18;  
Matches 20; Conservative 8; Mismatches 34; Indels 0; Gaps 0;

QY 30 LNEERAIDELKKQATIEDKEATTATEAASSDALEALADOTDALQSEEAHVVKADNAASDA 89  
DB 955 LREFDLOVNRKAEAEAMKRLSYISQKVSADSKTQQAERALGSAADAQRAKNKGAE 1014  
QY 90 LE 91  
DB 1015 LE 1016

RESULT 39  
US-08-317-450B-13  
Sequence 13, Application US/08317450B  
Patent No. 5660982  
GENERAL INFORMATION:  
APPLICANT: Tryggvason, Karl  
APPLICANT: Kallunki, Pekka  
APPLICANT: Pyke, Charles  
TITLE OF INVENTION: Laminin Chains: Diagnostic and  
TITLE OF INVENTION: Therapeutic Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER & ALLEGRETTI, LTD.  
STREET: Ten South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317.450B  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Chao, Mark  
REGISTRATION NUMBER: 37,293  
REFERENCE/DOCKET NUMBER: 94,778  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1193 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-450B-13

Query Match 8.2%; Score 64; DB 1; Length 1193;  
Best Local Similarity 32.3%; Pred. No. 20;

Job time : 27.6578 secs

THIS PAGE BLANK (USPTO)



C	1	108	13.9	442	9	AUI80004	AUI80004
C	2	108	13.9	570	9	AUI79963	AUI79963

Email: [naruse@biol.s.u-](mailto:naruse@biol.s.u-)

tokyo.ac.jp

tokyo.ac.jp







source 1. 517  
 /organism="Oryzias latipes"  
 /strain="HNI"  
 /db\_xref="taxon:8090"  
 /clone="Ole13.10d"  
 /clone\_lib="Medaka liver cDNA library (Ole) from HNI"  
 /tissue\_type="liver"  
 /dev\_stage="adult"  
 BASE COUNT 145 a 181 c 74 g 114 t 3 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.0289 Length: 517  
 Score: 97.00 Matches: 29  
 Percent Similarity: 41.44% Conservatives: 17  
 Best Local Similarity: 26.13% Mismatches: 65  
 Query Match: 12.48% Indels: 0  
 DB: 9 Gaps: 0

US-09-847-539A-6 (1-159) x AU179732 (1-517)

Qy 33 GluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThr 52  
 Db 465 GATGAGGCTGCTAGGATCTCTGAGGATCTGAGGCGCTGAGGATCTGAGGTTTGAAG 406  
 Qy 53 AlaIleGluAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 72  
 Db 405 GTACTGAGGCTGCTGAGGATCTCTGAGGATCTGAGGCGCTGAGGATCTGAGGTTTG 346  
 Qy 73 GlnSerGluGluAlaValValLysAlaAspAlaIleSerAspAlaLeuGluAla 92  
 Db 345 AAGGATAGTGGCTGCTGAGGATCTGAGGATCTGAGGCGCTGAGGATCTGAGGTTT 286  
 Qy 93 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAln 112  
 Db 285 TTGAAGGATAGTGGCTGCTGAGGATCTGAGGATCTGAGGCGCTGAGGATCTGAG 226  
 Qy 113 AlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysThr 132  
 Db 225 GTTTGAAGGATAGTGGCTGCTGAGGATCTGAGGATCTGAGGCGCTGAGGATCTGAG 166  
 Qy 133 LysAspThrLysProValValLysLysGluGlu 143  
 Db 165 GAGGATGCTGAGGCTGCTGAGGATCTGAGGATCTGAGGATCTGAGGATCTGAGGAT 133

RESULT 8  
 AU179899/c  
 LOCUS AU179899 380 bp mRNA linear EST 21-MAR-2001  
 DEFINITION AU179899 Medaka liver cDNA library (Ole) from HNI Oryzias latipes  
 cDNA clone Ole16.08a similar to gb|AB025967| Oryzias latipes mRNA  
 for chorogenin Hminor, complete cds, mRNA sequence.

ACCESSION AU179899  
 VERSION AU179899.1 GI:13428736  
 KEYWORDS EST.

SOURCE  
 ORGANISM

Oryzias latipes  
 Japanese medaka.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.  
 1 (bases 1 to 380)

REFERENCE  
 Naruse, K., Mitani, H. and Tanaka, M.  
 Medaka EST Project in University of Tokyo (2001)  
 Unpublished (2001)  
 CONTACT: Kiyoshi Naruse  
 Department of Biological Sciences  
 Graduate School of Science, University of Tokyo  
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan  
 Tel: 81-3-5841-4443  
 Fax: 81-3-5841-4410  
 Email: naruse@biol.s.u-tokyo.ac.jp  
 This clone was isolated from Medaka liver cDNA library (Ole).

JOURNAL  
 COMMENT  
 Location/Qualifiers

FEATURES

source 1. 380  
 /organism="Oryzias latipes"  
 /strain="HNI"  
 /db\_xref="taxon:8090"  
 /clone="Ole16.08a"  
 /clone\_lib="Medaka liver cDNA library (Ole) from HNI"  
 /tissue\_type="liver"  
 /dev\_stage="adult"  
 BASE COUNT 107 a 144 c 47 g 82 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.0263 Length: 380  
 Score: 96.00 Matches: 28  
 Percent Similarity: 41.94% Conservatives: 11  
 Best Local Similarity: 30.11% Mismatches: 54  
 Query Match: 12.36% Indels: 0  
 DB: 9 Gaps: 0

US-09-847-539A-6 (1-159) x AU179899 (1-380)

Qy 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysGln 43  
 Db 330 GCTGAGGATCTCTGAGGATCTCTGAGGATCTCTGAGGATCTCTGAGGATCTCTGAG 271  
 Qy 44 AlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGlu 63  
 Db 270 GCTGCTGAGGATCTCTGAGGATCTCTGAGGATCTCTGAGGATCTCTGAGGATCT 211  
 Qy 64 AlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAsp 83  
 Db 210 GAGGCTGCTGAGGATCTCTGAGGATCTCTGAGGATCTCTGAGGATCTCTGAGGAT 151  
 Qy 84 AsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 103  
 Db 150 ACTGAGGCTGATGAGGCTGCTGAGGATCTCTGAGGATCTCTGAGGATCTCTGAG 91  
 Qy 104 GluAlaGluValValGlnSerAspAlaAlaAlaSerAsp 116  
 Db 90 GATAGTGGCTGATGAGGCTGCTGAGGATCTCTGAGGATCTCTGAGGATCTCTGAG 52

RESULT 9  
 AZ895651

LOCUS  
 DEFINITION

AZ895651 567 bp DNA linear GSS 05-MAR-2001  
 RPCI-24-211J20.TJ RPCI-24 Mus musculus genomic clone RPCI-24-211J20  
 , DNA sequence.

ACCESSION AZ895651  
 VERSION AZ895651.1 GI:13214596  
 KEYWORDS GSS.

SOURCE  
 ORGANISM

Mus musculus  
 house mouse.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 567)

REFERENCE  
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
 Russell, D., de Jong, P. and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPCI-24  
 Unpublished (1999)  
 Other\_GSSs: RPCI-24-211J20.TV

JOURNAL  
 COMMENT

Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end  
 page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 211 row: J column: 20



```

KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Actinopterygii: Neopterygii: Teleostei: Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziatidae; Oryzias.
REFERENCE 1 (bases 1 to 394)
AUTHORS Naruse, K., Mitani, H. and Tanaka, M.
JOURNAL Medaka EST Project in University of Tokyo (2001)
COMMENT Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel.: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from Medaka liver cDNA library (OLE).
FEATURES
    source
    1..394
        /organism="Oryzias latipes"
        /strain="HNI"
        /db_xref="taxon:8090"
        /clone_lib="Medaka liver cDNA library (OLE) from HNI"
        /tissue_type="liver"
        /dev_stage="adult"
BASE COUNT 116 a 143 c 52 g 79 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 0.487 Length: 394
Score: 86.00 Matches: 25
Percent Similarity: 42.86% Conservative: 11
Best Local Similarity: 29.76% Mismatches: 48
Query Match: 11.07% Indels: 0
DB: 9 Gaps: 0
US-09-847-539a-6 (1-159) x AU179547 (1-394)
QY 33 GluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThr 52
Db 252 GATGAGCGCTCTAGGATTCCTGAGGATCTGAGCGGCTGAGGATTCCTGAGGTTTGAAG 193
QY 53 AlaIleGluAlaSerSerAspAlaLeuGluAlaLeuAlaSerGlnThrAspAlaLeu 72
Db 192 GGTACTGAGCGCTGAGGATTCCTNAGGATGNTGAGNCTGCTGAGGATTCCTGAGGTTTGT 133
QY 73 GlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAla 92
Db 132 AGGATAGTACGCGCTGAGGATTCCTGAGGATGTTGAGGCTGCTGAGGATTCCTGAGGTT 73
QY 93 LeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaGluAlaValGlnSerAspAsn 112
Db 72 TTGAAGGATACCTGAGGCTGCTGAGGATTCCTGAGGATCTGAGGCTGATGAGGCTGCTGAG 13
QY 113 AlaAlaSerAsp 116
Db 12 GTTTTGAAGGAT 1
RESULT 12
CNS0607Z/C 833 bp DNA linear GSS 04-JUL-2001
LOCUS T7 end of clone AV0AA004C03 of library AV0AA from strain CBS 379 of
DEFINITION Saccharomyces exiguus, genomic survey sequence.
ACCESSION AL407733
VERSION AL407733.1 GI:12173672
KEYWORDS GSS
SOURCE Saccharomyces exiguus.
ORGANISM Saccharomyces exiguus
Eukaryota: Fungi; Ascomycota: Saccharomycotina: Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

```

```

REFERENCE 1 (bases 1 to 833)
AUTHORS Bon, E., Neuveglise, C., Lepingie, A., Wincker, P., Artiguenave, F.,
Gaillardin, C. and Casaregola, S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 6.
JOURNAL Saccharomyces exiguus
MEDLINE PERS Lett. 487 (1), 42-46 (2000)
REFERENCE 20584716
REFERENCE 2 (bases 1 to 833)
AUTHORS Souciet, J.-L., Aigle, M., Artiguenave, F., Blandin, G.,
Boltin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Maupertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissbach, J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL PERS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
REFERENCE 3 (bases 1 to 833)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
    source
    1..833
        /organism="Saccharomyces exiguus"
        /strain="CBS 379"
        /db_xref="taxon:34358"
        /clone_lib="AV0AA004C03"
        /clone_lib="AV0AA"
        /note="end : 877"
        /note="similar to Saccharomyces cerevisiae ORF YBR067c [
        TIPI1 : esterase ]"
        /evidence="not_experimental"
        misc_feature
        <31..>258
        /note="similar to Saccharomyces cerevisiae ORF YOR010c [
        TIPI2 : cold shock induced protein ]"
        /evidence="not_experimental"
        misc_feature
        <31..>225
        /evidence="not_experimental"
        misc_feature
        <31..>225
        /note="similar to Saccharomyces cerevisiae ORF YER011w [
        TIPI1 : cold-shock induced protein of the Tirip, Tirip
        family ]"
        /evidence="not_experimental"
        misc_feature
        <34..>243
        /note="similar to Saccharomyces cerevisiae ORF YOR009w [
        similarity to Tirip and Tirip ]"
        /evidence="not_experimental"
BASE COUNT 165 a 191 c 128 g 332 t 17 others
ORIGIN
Alignment Scores:
Pred. No.: 1.24 Length: 833
Score: 86.00 Matches: 21
Percent Similarity: 48.81% Conservative: 20
Best Local Similarity: 25.00% Mismatches: 43
Query Match: 11.07% Indels: 0
DB: 12 Gaps: 0
US-09-847-539a-6 (1-159) x CNS0607Z (1-833)
QY 33 GluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThr 52

```











REFERENCE	4 (sites)																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
-----------	-----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



```

Query Match: 10.30% Indels: 0
DB: 11 Gaps: 0
US-09-847-539A-6 (1-159) x AK014987 (1-1158)

QY 33 GluGluAraAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThr 52
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 674 GAGGAGACCAAGGCTCTGACACACAGGAGGGCTCTCAGAAAGGAAAGCTCTG 733
QY 53 AlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 72
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 734 ATGAGGATGGAGTGGCGGCGGAGCTGGCGAGCGGCTCTACCAAGGAGGGGCTGCTCTG 793
QY 73 GlnSerGluGluAlaValValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAla 92
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 794 GAGATGGAAGACAGGCTCTGTGTGAGGAGGACCGAGCTCTCCGGGAGGAGCAAGCA 853
QY 93 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu 104
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 854 CTCAGAGAGAGATGGGCTCTCCAAAGATGAAGAA 889

RESULT 23
TA249H110/c 556 bp DNA linear GSS 13-DEC-2000
LOCUS TA249H110/c 556 bp DNA linear GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 249h11, reverse sequence,
genomic survey sequence.
ACCESSION AL482135
VERSION AL482135.1 GI:11848140
KEYWORDS GSS
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 556)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/Tbrucei/.

FEATURES
source
1..556
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="249h11"
BASE COUNT 88 a 186 c 136 g 146 t
ORIGIN

Alignment Scores: 5.6 Length: 556
Pred. No.: 79.00 Matches: 24
Score: 42.50% Conservative: 27
Percent Similarity: 20.00% Mismatches: 69
Best Local Similarity: 10.17% Indels: 0
Query Match: 12 Gaps: 0
DB: 12

US-09-847-539A-6 (1-159) x TA249H110 (1-556)

QY 38 AspGluLeuLysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAla 57
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 504 GAACAAATCAGGCGCTCTATTCAAAATGACGACAGCGTAGACAGCGGAGCAAGCG 445
QY 58 SerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 77
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 444 GTGGAGCAGCGCCGACGAGCCCTTGAGGCGGCAAGAAATGACAGCATCCCGCTAAGGAT 385
QY 78 AlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThr 97
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 384 GCGTCCGAACTCGGATCGGATCTCAAGATGTCAGAGCGCGCCATCCGCGAGAAGAGGATCGC 325
QY 98 AspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAla 117
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 324 AGGGACATGGACAACCAAGCGCGAGAGGAGCGCGCCATCCGCGAGAAGAGGATCGC 265
QY 118 TrpGluLysAlaAlaThrProLeuAlaLeuAspValLysLysThrLysAspThrLysPro 137
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 GAGAAGAACTCTTCGAGATTCCTGAGCCTGAAAGAGGCTTCGGAAGACCGCACAAGAG 205
QY 138 ValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSer 157
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 204 GTGGCTGTGAGCTGAGAGCAGGAGCGACGCGCGCTTTCTCTCGCTGGGAGCGAGT 145

RESULT 24
AQ875889/c 797 bp DNA linear GSS 08-NOV-1999
LOCUS AQ875889/c 797 bp DNA linear GSS 08-NOV-1999
DEFINITION V130F6 mTn-3xHA/lacZ Insertion Library, strain Y2278 Saccharomyces
cerevisiae genomic 5', DNA sequence.
ACCESSION AQ875889
VERSION AQ875889.1 GI:6288133
KEYWORDS GSS.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 797)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
Desjardes, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R.,
umansky, L., Heidman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
Gene Disruption
Unpublished (1999)
Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mTn-3xHA/lacZ insertion.
Seq primer: GGCCTCTCTCTTCTTGGAGTAGTAC
Class: transposon-tagged.
Location/Qualifiers
1..797
/organism="Saccharomyces cerevisiae"
/strain="Y2278 - S288C background, cir(0) rho(0)"
/db_xref="taxon:4932"
/clone_lib="mtn-3xHA/lacZ Insertion Library, strain Y2278"
/lab_host="E. coli"
/note="Vector: pHS56-Sal; A yeast genomic DNA library
without 2 micron or mitochondrial DNA was prepared in
pHS56-Sal; genomic DNA was size-fractionated (DNA of
roughly 2-3 kb in length) prior to cloning. This library
was subsequently mutagenized with a mtn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT 206 a 175 c 209 g 207 t
ORIGIN

Alignment Scores: 8.76 Length: 797
Pred. No.:

```







```
/clone_lib="cosmid 152F05"
/clone="152F05BE3"
BASE COUNT      55 a      122 c      72 g      84 t      5 others
ORIGIN

Alignment Scores:
  Pred. No.:      4.01      Length:      338
  Score:          78.00      Matches:      21
  Percent Similarity: 48.78%      Conservative: 19
  Best Local Similarity: 25.61%      Mismatches: 42
  Query Match:      10.04%      Indels:      0
  DB:              12      Gaps:      0

US-09-847-539A-6 (1-159) x FR0042547 (1-338)
QY 31 ArgAsnGluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAla 50
Db 295 GCACAGATGGAGCGAGGCTGCCAGCGCCCTCAAGCAGAGGAGGAGGAGGAGGAG 236
QY 51 ThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 70
Db 235 TGGGTCAAGAGCTCAGCAGAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 176
QY 71 AlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeu 90
Db 175 GCTTTGGAAGGCAGAGCGGAGGTGTGGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAG 116
QY 91 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValGlnSer 110
Db 115 GAGGCGAGCGGTGTCAGTCAGGAGGCTCTGGAGAGGAGGAGGAGGAGGAGGAGGAG 56
QY 111 AspAsn 112
Db 55 GAGAGC 50

RESULT 29
AO875428/c
LOCUS      AO875428      494 bp      DNA      linear      GSS      08-NOV-1999
DEFINITION cerevisiae genomic 5', DNA sequence.
ACCESSION      AO875428.1      GI:6287672
VERSION      GSS.
KEYWORDS      baker's yeast.
SOURCE      Saccharomyces cerevisiae
ORGANISM      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
REFERENCE      Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
AUTHORS      1 (bases 1 to 494)
      Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
      desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R.,
      Umansky, L., Heideman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
      Hager, K., Miller, P., Roeder, G.S., and Snyder, M.
      Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
      Gene Disruption
      Unpublished (1999)
      Contact: Kumar A
      Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
      Yale University
      P.O. Box 208103, New Haven, CT 06520-8103, USA
      Tel: 203 432 9949
      Fax: 203 432 6161
      Email: anuj.kumar@yale.edu
      te of mTn-3xHA/lacZ insertion.
      Seq primer: GCCTCTCTTTTGGCAAGTAC
      Class: transposon-tagged.
      Location/Qualifiers
        1..494
          /organism="Saccharomyces cerevisiae"
          /strain="Y2278 - S288C background, cir(0) rho(0)"
          /db_xref="taxon:4932"
          /clone_lib="mTn-3xHA/lacZ Insertion Library, strain Y2278"
          /lab_host="E. coli"
          /note="Vector: pHS56-Sal; A yeast genomic DNA library"

FEATURES
      source
        1..498
          /organism="Meloidogyne arenaria"
          /db_xref="taxon:6304"
          /clone_lib="Meloidogyne arenaria egg pAMP1 v1 Chiapelli"

without 2 micron or mitochondrial DNA was prepared in
pHS56-Sal; genomic DNA was size-fractionated (DNA of
roughly 2-3 kb in length) prior to cloning. This library
was subsequently mutagenized with a mTn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance.

BASE COUNT      133 a      96 c      92 g      173 t
ORIGIN

Alignment Scores:
  Pred. No.:      6.44      Length:      494
  Score:          78.00      Matches:      19
  Percent Similarity: 56.86%      Conservative: 10
  Best Local Similarity: 37.25%      Mismatches: 22
  Query Match:      10.04%      Indels:      0
  DB:              12      Gaps:      0

US-09-847-539A-6 (1-159) x AO875428 (1-494)
QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
Db 258 AANGCTGAACCTCCGCCCAAGATGTCCAAACAAGTTGGAGAACCAAGAACTTTG 199
QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
Db 198 CAAACAAGGCGCAAGAGTAAAGGAACAAGCTGAAGCTTCTATCGACAACCTATAAAT 139
QY 121 AlaAlaThrProIleAlaLeuAspValLysLys 131
Db 138 GAAGCTACTCCAGAAGCTGAACAGGTGAAGAAG 106

RESULT 30
BI863069/c
LOCUS      BI863069      498 bp      mRNA      linear      EST      10-OCT-2001
DEFINITION rmi6b06.y1 Meloidogyne arenaria egg pAMP1 v1 Chiapelli McCarter
ACCESSION      BI863069
VERSION      BI863069.1      GI:16004859
KEYWORDS      EST.
SOURCE      Meloidogyne arenaria.
ORGANISM      Meloidogyne arenaria
      Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchinea;
      Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE      1 (bases 1 to 498)
AUTHORS      McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
      Dante, M., Marra, M., Hillier, B., Kucaba, T., Theising, B., Bowers, Y.,
      Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V., R.,
      Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe
      M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
      Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and
      Wilson, R.
      The Washington Univ. Nematode EST Project, 1999
      Unpublished (1999)
      Contact: McCarter JP
      The Washington Univ. Nematode EST Project, 1999
      Washington University School of Medicine
      4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
      Tel: 314 286 1800
      Fax: 314 286 1810
      Email: est@wustl.edu
      The library was constructed by Brandi Chiapelli and Dr. James
      McCarter (bchiapell@wustl.edu & jmccartee@wustl.edu) at
      Washington University, St. Louis. DNA Sequencing by: Washington
      University Genome Sequencing Center St. Louis. Nematodes were
      provided by Dr. David Bird and Daniel Snyder of North Carolina
      State University.
      Seq primer: -40RP from Gibco
      High quality sequence stop: 415.
      Location/Qualifiers
        1..498
          /organism="Meloidogyne arenaria"
          /db_xref="taxon:6304"
          /clone_lib="Meloidogyne arenaria egg pAMP1 v1 Chiapelli"
```



```

/db_xref="taxon:7091"
/clone="ws00597"
/clone_lib="Bombyx mori p50(Daizo)"
BASE COUNT      243 a 177 c 220 g 113 t
ORIGIN

Alignment Scores:      Length: 753
Pred. No.:            10.9  Matches: 24
Score:                78.00  Conservative: 23
Percent Similarity:   44.34%  Mismatches: 59
Best Local Similarity: 22.64%  Indels: 0
Query Match:         10.04%   Gaps: 0
DB:

US-09-847-539A-6 (1-159) x AU003803 (1-753)
Qy 37 IleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 56
Db 291 CTGATCAGCTGCAGCGCGCAAGCTTAAAGAGTTCAAGTGTGGAACCTCAGCAAGCGCAC 350
Qy 57 AlaSerSerAspAlaLeuAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu 76
Db 351 GCCCAATGAGATTACAGACTTAAAGAGTTCAAGTGTGGAACCTCAGCAAGCGCAC 410
Qy 77 AlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGln 96
Db 411 AAGCAAGCGATCCCAAGCCACGATGACGTCGCCGCCCTCCCAAGAGAGGTAGCGCGGCC-470
Qy 97 ThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAsp 116
Db 471 GAGGAGAGCATCAACAGCATCCAGCAGGAGCTGTCGCCGCCAGTGCAGCAAGCGCGAG 530
Qy 117 AlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysThrLysAspThrLys 136
Db 531 GCCAGGACTTGGCCCGCAGAGCCCAAGAGATATCCCGACAAAGCTTCTGTGAGCGCGCC 590
Qy 137 ProValValLysLysGlu 142
Db 591 CACGAATCAAGAGAAA 608

RESULT 33
AJ395963
LOCUS      AJ395963      755 bp      mRNA      linear      EST 25-JAN-2001
DEFINITION AJ395963 dkfz426 Gallus gallus cDNA clone 24ml7rl, mRNA sequence.
ACCESSION  AJ395963
VERSION     AJ395963.1 GI:7127417
KEYWORDS    EST.
SOURCE      chicken.
ORGANISM    Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 755)
AUTHORS    Adrakhmanov,I., Lodygin,D., Geroth,P., Arakawa,H., Law,A., Plachy
            J., Korn,B. and Buerstedde,J.M.
TITLE      A large database of chickenursal ESTs as a resource for the
            analysis of vertebrate gene function
JOURNAL    Genome Res. 10 (12), 2062-2069 (2000)
MEDLINE    20568495
COMMENT    Contact: Buerstedde JM
            Cellular Immunology
            Heinrich-Pette-Institute
            Martinistr. 52, 20251 Hamburg, Germany
            Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
            Location/Qualifiers
            1..755
               /organism="Gallus gallus"
               /strain="CB"
               /db_xref="taxon:9031"
               /clone="24ml7rl"
               /clone_lib="dkfz426"
               /tissue_type="Bursa of Fabricius"
BASE COUNT      315 a 124 c 185 g 131 t

```

---

```

ORIGIN

Alignment Scores:      Length: 755
Pred. No.:            10.9  Matches: 22
Score:                78.00  Conservative: 18
Percent Similarity:   47.06%  Mismatches: 45
Best Local Similarity: 25.88%  Indels: 0
Query Match:         10.04%   Gaps: 0
DB:

US-09-847-539A-6 (1-159) x AJ395963 (1-755)
Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLys 81
Db 214 GTTCAGCATTAGTTCTGGAACCTTGAACAGCTGAACAGCAAGCAAGCTCATATAAGCAA 273
Qy 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101
Db 274 CAGAGTCAGGCGAGCACCAACCAATCCATCCCTTAAAGGACCAAGTCAGTCTTTGGAA 333
Qy 102 SerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAla 121
Db 334 GCTGAGCAGTGAAGACAGGGAATCTTAAAGAACGCGAGAGATGAGCTGTCAGTGAA 393
Qy 122 AlaThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLysLys 141
Db 394 AAGGATTATGGCAGACGACGAACTAAAGATATTAAAGCCAATCTGCAAGATAGAGAA 453
Qy 142 GluGluArgGlnAsn 146
Db 454 TACAGAGAGCAAAAT 468

RESULT 34
CNS07CXJ/c
LOCUS      CNS07CXJ      968 bp      DNA      linear      GSS 08-JUL-2001
DEFINITION T7 end of clone BD0AA007F08 of library BD0AA from strain CBS 94 of
            Candida tropicalis, genomic survey sequence.
ACCESSION  AL439757
VERSION     AL439757.1 GI:12223169
KEYWORDS    GSS.
SOURCE      Candida tropicalis.
ORGANISM    Candida tropicalis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1 (bases 1 to 968)
AUTHORS    Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
            Dujon,B.
TITLE      Genomic exploration of the hemiascomycetous yeasts: 16. Candida
            tropicalis
JOURNAL    FEBS Lett. 487 (1), 91-94 (2000)
MEDLINE    20584726
REFERENCE   2 (bases 1 to 968)
AUTHORS    Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhata,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissbach,J.
TITLE      Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
JOURNAL    FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE    20584711
REFERENCE   3 (bases 1 to 968)
AUTHORS    Direct Submission
TITLE      Direct Submission
JOURNAL
COMMENT    This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia

```





US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: candersn@pw.usda.gov  
Sequence have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
seq primer: Stratagene SK primer.

FEATURES  
source

```

Source
i: .003
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE2336_F03_L06"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid:
Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give pBluescript
phagemids in the TJ close lab (Choi, Close, Fenton) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."
```

BASE COUNT	135 a	262 c	188 g	78 t
ORIGIN				

Alignment Scores:	
Pred. No.:	12.4
Score:	77.00
Percent Similarity:	39.32%
Best Local Similarity:	27.35%
Query Match:	9.91%
DB:	10
	Length: 663
	Matches: 32
	Conservative: 14
	Mismatches: 70
	Indels: 1
	Gaps: 0

```

US-09-847-539A-6 (1-159) x BF483682 (1-663)

```

RESULT	39
AU003869	
LOCUS	
DEFINITION	Bombyx mori p50(Daizo) Bombyx mori cDNA clone ws00687, mRNA sequence.
ACCESSION	AU003869
VERSION	AU003869.1 GI:4161240
KEYWORDS	
DATA FILE	
FILE SIZE	
MIME TYPE	
MD5 CHECKSUM	
SHA-1 CHECKSUM	
SHA-256 CHECKSUM	
EST ID	EST 19-JAN-1999
LINEAR	linear
ORIGIN	
FEATURES	
CDS	
ORF	
ORF1	
ORF2	
ORF3	
ORF4	
ORF5	
ORF6	
ORF7	
ORF8	
ORF9	
ORF10	
ORF11	
ORF12	
ORF13	
ORF14	
ORF15	
ORF16	
ORF17	
ORF18	
ORF19	
ORF20	
ORF21	
ORF22	
ORF23	
ORF24	
ORF25	
ORF26	
ORF27	
ORF28	
ORF29	
ORF30	
ORF31	
ORF32	
ORF33	
ORF34	
ORF35	
ORF36	
ORF37	
ORF38	
ORF39	
ORF40	
ORF41	
ORF42	
ORF43	
ORF44	
ORF45	
ORF46	
ORF47	
ORF48	
ORF49	
ORF50	
ORF51	
ORF52	
ORF53	
ORF54	
ORF55	
ORF56	
ORF57	
ORF58	
ORF59	
ORF60	
ORF61	
ORF62	
ORF63	
ORF64	
ORF65	
ORF66	
ORF67	
ORF68	
ORF69	
ORF70	
ORF71	
ORF72	
ORF73	
ORF74	
ORF75	
ORF76	
ORF77	
ORF78	
ORF79	
ORF80	
ORF81	
ORF82	
ORF83	
ORF84	
ORF85	
ORF86	
ORF87	
ORF88	
ORF89	
ORF90	
ORF91	
ORF92	
ORF93	
ORF94	
ORF95	
ORF96	
ORF97	
ORF98	
ORF99	
ORF100	
ORF101	
ORF102	
ORF103	
ORF104	
ORF105	
ORF106	
ORF107	
ORF108	
ORF109	
ORF110	
ORF111	
ORF112	
ORF113	
ORF114	
ORF115	
ORF116	
ORF117	
ORF118	
ORF119	
ORF120	
ORF121	
ORF122	
ORF123	
ORF124	
ORF125	
ORF126	
ORF127	
ORF128	
ORF129	
ORF130	
ORF131	
ORF132	
ORF133	
ORF134	
ORF135	
ORF136	
ORF137	
ORF138	
ORF139	
ORF140	
ORF141	
ORF142	
ORF143	
ORF144	
ORF145	
ORF146	
ORF147	
ORF148	
ORF149	
ORF150	
ORF151	
ORF152	
ORF153	
ORF154	
ORF155	
ORF156	
ORF157	
ORF158	
ORF159	
ORF160	
ORF161	
ORF162	
ORF163	
ORF164	
ORF165	
ORF166	
ORF167	
ORF168	
ORF169	
ORF170	
ORF171	
ORF172	
ORF173	
ORF174	
ORF175	
ORF176	
ORF177	
ORF178	
ORF179	
ORF180	
ORF181	
ORF182	
ORF183	
ORF184	
ORF185	
ORF186	
ORF187	
ORF188	
ORF189	
ORF190	
ORF191	
ORF192	
ORF193	
ORF194	
ORF195	
ORF196	
ORF197	
ORF198	
ORF199	
ORF200	

**KEYWORDS**  
**SOURCE**  
**ORGANISM**

EST.  
domestic silkworm.  
Bombyx mori  
ORGANISM  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia  
1 (bases 1 to 675)  
Bombycoidea; Bombycidae; Bombyx.  
REFERENCE  
Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.  
Establishment of cDNA database of Bombyx mori  
Unpublished (1999)  
JOURNAL

COMMENT

FEATURES	source	BASE COUNT	ORIGIN	Alignment :	Pred. NO.:	Percent Ship	Best Local	Query Match	DB:
----------	--------	------------	--------	-------------	------------	--------------	------------	-------------	-----

US-09-847-1

Qy	37	291	57	351	77	411	97	471	117	531	137	591	RESULT 40	AZ948693	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
----	----	-----	----	-----	----	-----	----	-----	-----	-----	-----	-----	-----------	----------	-------	------------	-----------	---------	----------	--------	----------	-----------

## AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0211 row: 0 column: 18

Seq primer: CACACAGGAAACGCTATGACC

Class: plasmid ends

High quality sequence stop: 723.

Location/Qualifiers

1..723

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0211018"

/clone\_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (female) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adaptored DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (g1473211419b1AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adaptored mouse DNA was annealed to

adaptored vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 245 a 133 c 254 g 91 t

ORIGIN

Alignment Scores:

Pred. No.: 13.8 Length: 723

Score: 77.00 Matches: 25

Percent Similarity: 46.15% Conservative: 17

Best Local Similarity: 27.47% Mismatches: 49

Query Match: 9.91% Indels: 0

DB: 12 Gaps: 0

US-09-847-539A-6 (1-159) x AZ948693 (1-723)

QY 31 ArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGluAla 50

Db 6 CGCGAGCGCAGAGGAGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGA 65

QY 51 ThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 70

Db 66 GAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 125

QY 71 AlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeu 90

Db 126 GCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 185

QY 91 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSer 110

Db 186 GCAGAGGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGG 245

QY 111 AspAsnAlaAlaSerAspAlaTrpGluLysAla 121

Db 246 GCAGAGGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCA 278

Search completed: October 13, 2002, 03:33:06

Job time : 1894.74 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: October 13, 2002, 01:12:51 : Search time 22.9572 Seconds  
(without alignments)  
268.169 Million cell updates/sec

Title: US-09-847-539A-6  
Perfect score: 777  
Sequence: 1 VDSPIEQPRIIPNGGTLN.....KKEERQNVNTLPTTGESNP 159

Scoring table: BLOSUM62  
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	155	19.9	448	1 SPGL_STRSP	P06654 streptococ
2	155	19.9	593	1 SPGL_STRSP	P19909 streptococ
3	99	12.7	774	1 STFL_LAMBD	P03764 bacterioph
4	88	11.3	407	1 IE68_HSVSA	Q01042 herpesvirs
5	85	10.9	1120	1 STFR_ECOLI	P76072 escherichia
6	80	10.3	269	1 TIR3_YEAST	P40552 saccharomyc
7	80	10.3	347	1 TOLA_PSEAE	P50600 pseudomonas
8	79	10.2	389	1 YM82_YEAST	Q04951 saccharomyc
9	79	10.2	516	1 P54_ENTFC	P13692 enterococcu
10	78	10.0	771	1 CALD_CHICK	P12957 gallus gall
11	76	9.8	490	1 CE05_ECOLI	Q47500 escherichia
12	76	9.8	490	1 CE10_ECOLI	Q47125 escherichia
13	76	9.8	1290	1 XPCX_XENLA	P50532 xenopus lae
14	75	9.7	406	1 SR40_YEAST	P32583 saccharomyc
15	75	9.7	3210	1 CENF_HUMAN	P49454 homo sapien
16	73	9.4	1189	1 SCIL_CHICK	Q90988 gallus gall
17	73	9.4	1278	1 DINA_HUMAN	Q44203 homo sapien
18	73	9.4	1396	1 VLTF_BPT5	P13390 bacterioph
19	72	9.3	464	1 AK15_RAT	P24587 rattus norv
20	72	9.3	488	1 CVL2_BOVIN	Q28092 bos taurus
21	72	9.3	2230	1 GOG4_HUMAN	Q13439 homo sapien
22	71	9.1	217	1 GPPE_HALN1	Q9hrv0 halobacteri
23	71	9.1	439	1 DDXT_HUMAN	Q15320 homo sapien
24	71	9.1	1238	1 SBCC_RHOCA	O68032 rhodobacter
25	70	9.0	211	1 T2D9_HUMAN	Q15544 homo sapien
26	70	9.0	211	1 TUB8_SOLTU	P33191 solanum tub
27	70	9.0	421	1 TOLA_ECOLI	P19934 escherichia
28	70	9.0	433	1 HTR2_HALVA	P42258 halocaula
29	70	9.0	857	1 NPM_CHICK	P16053 gallus gall
30	70	9.0	1391	1 MST2_DROHY	Q08696 drosophila
31	69	8.9	198	1 HBHA_MYCTU	Q11142 mycobacteri
32	69	8.9	500	1 GAR2_SCHPO	P41891 schizosacch
33	69	8.9	889	1 HS97_STRPU	Q06068 strongyloce

34 69 8.9 1131 1 AC15\_MOUSE  
35 69 8.9 1281 1 DYNA\_MOUSE  
36 69 8.9 1423 1 FRUA\_STRMU  
37 69 8.9 1528 1 SPAA\_STRDO  
38 68 8.8 232 1 YJFJ\_ECOLI  
39 68 8.8 643 1 DP3X\_ECOLI  
40 67 8.6 654 1 DNAK\_CHLMU  
41 67 8.6 756 1 Y328\_MYCGE  
42 67 8.6 783 1 HTR2\_HALN1  
43 67 8.6 784 1 HTR2\_HALSA  
44 67 8.6 1399 1 RPOC\_PSEAE  
45 67 8.6 1407 1 RPOC\_BUCAI

P35601 mus musculu  
O08788 mus musculu  
Q03174 streptococ  
P21979 streptococ  
P39292 escherichia  
P06710 escherichia  
P56836 chlamydia m  
O49419 mycoplasma  
Q9hp81 halobacteri  
P71410 halobacteri  
Q9hwc9 pseudomonas  
P57145 buchnera ap

## ALIGNMENTS

RESULT 1  
SPGL\_STRSP  
ID SPGL\_STRSP STANDARD; PRT; 448 AA.  
AC P06654;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Immunoglobulin G binding protein G precursor (IGG binding protein G).  
GN SPG.  
OS Streptococcus sp. (Lancefield group G).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1306;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=863041178; PubMed=3745123;  
RA Fahnestock S.R., Alexander P., Nagle J., Filpula D.;  
RT "Gene for an immunoglobulin-binding protein from a group G streptococcus.";  
RL J. Bacteriol. 167:870-880(1986).  
RN [2]  
RX X-RAY CRYSTALLOGRAPHY (2.07 ANGSTROMS) OF 228-282.  
RA MEDLINE=94213848; PubMed=8161530;  
RT Gallagher T., Alexander P., Bryan P., Gilliland G.L.;  
RT "Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR.";  
RN [3]  
RX Biochemistry 33:4721-4729(1994).  
RX X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 293-351.  
RA MEDLINE=95055731; PubMed=7966308;  
RT Derrick J.P., Wigley D.B.;  
RT "The third IgG-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab.";  
RL J. Mol. Biol. 243:906-918(1994).  
RN [4]  
RX X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 297-352.  
RA MEDLINE=95308043; PubMed=7788293;  
RT Sauer-Eriksson A.E., Kleywegt G.J., Uhlen M., Jones T.A.;  
RT "Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG.";  
RN [5]  
RX X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 292-352.  
RA Butterworth S., Lanzin V.S., Wigley D.B., Derrick J.P., Wilson K.S.;  
RT Submitted (APR-1997) to the PDB data bank.  
RN [6]  
RX STRUCTURE BY NMR OF 298-351.  
RA MEDLINE=91335209; PubMed=1871600;  
RT Gronenborn A.M., Filpula D.R., Essig N.Z., Achari A., Whitlow M., Wingfield P.T., Clore G.M.;  
RT "A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G.";  
RL Science 253:657-661(1991).  
CC -!- FUNCTION: BINDS TO THE CONSTANT FC REGION OF IGG WITH HIGH AFFINITY.

```
CC CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: M13825; AAA03664.1; -
CC DR PIR: A24496; A24496.
CC DR PDB: 1PGA; 30-APR-94.
CC DR PDB: 1PGX; 15-JUL-92.
CC DR PDB: 1IGC; 03-JUN-95.
CC DR PDB: 1IGD; 01-NOV-94.
CC DR PDB: 2IGD; 29-JUL-98.
CC DR PDB: 1FGC; 20-JUL-95.
CC DR PDB: 1GB1; 15-APR-93.
CC DR InterPro: IPR002988; GA.
CC DR InterPro: IPR001899; Gram_pos_anchor.
CC DR InterPro: IPR000724; IgG_bind_B.
CC DR Pfam: PF01468; GA; 2.
CC DR Pfam: PF00746; Gram_pos_anchor; 1.
CC DR Pfam: PF01378; IgG_binding_B; 2.
CC DR PRINTS: PR00015; GPOSANCHOR.
CC DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
CC DR IgG-binding protein; Repeat; Transmembrane; Cell wall; Signal;
CC 3D-structure.
CC FT SIGNAL 1 33 IMMUNOGLOBULIN G BINDING PROTEIN G.
CC FT CHAIN 34 448 EXTRACELLULAR.
CC FT DOMAIN 423 443
CC FT TRANSMEM 444 448
CC FT DOMAIN 69 111 CYTOPLASMIC.
CC FT DOMAIN 104 215 ALA-RICH.
CC FT REPEAT 104 140 2 X 37 AA REPEATS.
CC FT REPEAT 179 215 1-1.
CC FT REPEAT 228 352 1-2.
CC FT REPEAT 228 352 2 X 55 AA REPEATS.
CC FT REPEAT 298 352 2-1.
CC FT REPEAT 386 410 2-2.
CC FT DOMAIN 386 410 5 X 5 AA REPEATS OF [DE]-D-A-K-K.
CC FT DOMAIN 414 419 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
CC FT PROTEINS.
CC SQ SEQUENCE 448 AA; 47567 MW; A0759060C8F956CA CRC64;
Query Match 19.9%; Score 155; DB 1; Length 448;
Best Local Similarity 56.1%; Pred. No. 1.3e-09;
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;
QY 1 VDSPIQPIIPNGTTLNLGNAPKLAIRNEERAIDELKKQAIEDKATTAIEAASD 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34 VDSPIEDTPIIRNGBELNLLGNSETTLALRNEESATADLTAAAVADTVAAANAGAA 93
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 ALEALA 66
| | | |
Db 94 AWEAAA 99
RESULT 2
SPG2_STRSP STANDARD; PRT; 593 AA.
AC P19909;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Immunoglobulin G binding protein G precursor (IgG binding protein G).
GN SPG.
OS Streptococcus sp. (strain G148).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
```

```
OC Streptococcus.
RN NCBI_TaxID=1306;
RP SEQUENCE FROM N.A.
RC STRAIN=G148;
RX MEDLINE=88029445; PubMed=3665928;
RA Olsson A., Eliasson M., Guss B., Nilsson B., Hellman U.,
Lindberg M., Uhlen M.;
RT "Structure and evolution of the repetitive gene encoding
streptococcal protein G.";
RL Eur. J. Biochem. 168:319-324(1987).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=GX7805;
RX MEDLINE=88015586; PubMed=3658689;
RA Filipula D., Alexander P., Fahnestock S.R.;
RT "Nucleotide sequence of the protein G gene from Streptococcus GX7805,
and comparison to previously reported sequences.";
RL Nucleic Acids Res. 15:7210-7210(1987).
RN [3]
RC SEQUENCE OF 114-593 FROM N.A.
RP STRAIN=G148;
RX MEDLINE=863300657; PubMed=3017704;
RA Guss B., Eliasson M., Olsson A., Uhlen M., Frej A.-K., Joernvall H.,
Flock J.-I., Lindberg M.;
RT "Structure of the IgG-binding regions of streptococcal protein G.";
RL EMBO J. 5:1567-1575(1986).
RN [4]
RP STRUCTURE BY NMR OF 371-427.
RC STRAIN=G148;
RX MEDLINE=98290449; PubMed=9628485;
RA Malakuskas S.M., Mayo S.L.;
RT "Design, structure and stability of a hyperthermophilic protein
variant.";
RL Nat. Struct. Biol. 5:470-475(1998).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: X06173; CAA29540.1; -
CC DR EMBL: Y00428; CAA68489.1; -
CC DR PIR: S00128; S00128.
CC DR PIR: A26314; A26314.
CC DR PDB: 1GB4; 22-JUL-98.
CC DR InterPro: IPR002988; GA.
CC DR InterPro: IPR001899; Gram_pos_anchor.
CC DR InterPro: IPR000724; IgG_bind_B.
CC DR Pfam: PF01468; GA; 3.
CC DR Pfam: PF00746; Gram_pos_anchor; 1.
CC DR Pfam: PF01378; IgG_binding_B; 3.
CC DR PRINTS: PR00015; GPOSANCHOR.
CC DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
CC IgG-binding protein; Repeat; Transmembrane; Cell wall; Signal;
CC 3D-structure.
CC FT SIGNAL 1 33 IMMUNOGLOBULIN G BINDING PROTEIN G.
CC FT CHAIN 34 593 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 34 567 MEMBRANE ANCHOR.
CC FT TRANSMEM 568 588 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 589 593 ALA-RICH.
CC FT DOMAIN 59 111 3 X 37 AA REPEATS.
CC FT REPEAT 104 240 1-1.
CC FT REPEAT 179 215 1-2.
CC FT REPEAT 254 290 1-3.
```

```
FT DOMAIN 303 427 2 X 55 AA REPEATS.
FT REPEAT 303 357 2-1.
FT REPEAT 373 427 2-2.
FT DOMAIN 531 555 5 X 5 AA REPEATS OF [DE]-D-A-K-K.
FT DOMAIN 559 564 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
                          PROTEINS.
SQ SEQUENCE 593 AA; 63291 MW; 048BAA760D5B2920 CRC64;

Query Match 19.9%; Score 155; DB 1; Length 593;
Best Local Similarity 55.1%; Pred. No. 1.7e-09;
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 VDSPIEPIIPNGCTLTNLGNAPKALNEERAIDELKQAIEDKQATTAIEAASSD 60
      ||||| || ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34 VDSPIEDTPIIRNGGELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGAA 93

QY 61 ALEALA 66
      |||||
Db 94 AWEAAA 99

RESULT 3
STF_LAMBD
ID STF_LAMBD STANDARD; PRT; 774 AA.
AC P03764; P03745;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Side tail fiber protein.
GN STF.
OS Bacteriophage lambda.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10710;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83189071; PubMed=6221115;
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT "Nucleotide sequence of bacteriophage lambda DNA.";
RL J. Mol. Biol. 162:729-773(1982).
RN [2]
RP IDENTIFICATION AS STF.
RX MEDLINE=92165720; PubMed=1531648;
RA Haggard-Ljungquist E., Halling C., Calendar R.;
RT "DNA sequences of the tail fiber genes of bacteriophage P2: evidence
RT for horizontal transfer of tail fiber genes among unrelated
RT bacteriophages.";
RL J. Bacteriol. 174:1462-1477(1992).
RN [3]
RP RECONSTRUCTION OF STF.
RX MEDLINE=93068310; PubMed=1439823;
RA Hendrix R.W., Duda R.L.;
RT "Bacteriophage lambda P2a: not the mother of all lambda phages.";
RL Science 258:1145-1148(1992).
CC -1- MISCELLANEOUS: The common laboratory strain of bacteriophage
CC lambda; lambda P2a; carries a frameshift mutation relative to Ur-
CC lambda, the original isolate. The Ur-lambda virions have thin,
CC jointed tail fibers (side tail fibers) that are absent from lambda
CC wild type. Relative to lambda P2a, Ur-lambda has expanded
CC receptor specificity and adsorbs to E.coli cells more rapidly.
CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC CORRECTED IN POSITION TO 396 TO RECREATE THE ORIGINAL STF PROTEIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; J02459; AAA96555.1; ALT_FRAME.
```

```
DR EMBL; J02459; AAA96557.1; ALT_FRAME.
DR PIR; A04389; QXBP1L.
DR PIR; A04370; QXBP2L.
DR InterPro; IPR004089; Chemotaxis_transducer.
KW Fiber protein.
SQ SEQUENCE 774 AA; 77527 MW; CDD1DF85E919123B CRC64;

Query Match 12.7%; Score 99; DB 1; Length 774;
Best Local Similarity 32.9%; Pred. No. 0.0033;
Matches 26; Conservative 16; Mismatches 37; Indels 0; Gaps 0;

QY 50 ATTAIEAASSDALEALADOTDALQSEEAAYVKADNAASDALEALADOTDALQSEEAAYVQ 109
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 AESSKNAATSAQAKTSETNAASQQAATSAATKASEAATSARDAVASKEAAKSS 249

QY 110 SDNAASDAWEKAATPIALD 128
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 250 ETNASSAGRAASSATAAE 268

RESULT 4
IE68_HSVSA
ID IE68_HSVSA STANDARD; PRT; 407 AA.
AC Q01042;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Immediate-early protein.
GN 73 OR ECUFI.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230228; PubMed=1314457;
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT organization between HVS and Epstein-Barr virus.";
RL Virology 188:296-310(1992).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 AND
CC HSV-2 IE-68 (US1), EBV-1 65, EBV-4 (ORF4), PRV RSP40, AND VZV 63.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X64346; CAA45696.1; -.
DR EMBL; M86409; AAA6149.1; -.
DR EMBL; S76368; AAB21116.1; -.
DR PIR; G36813; EDBEQ3.
DR HSSP; S20244; S20244.
DR HSSP; P53041; 1A17.
KW Early protein.
FT DOMAIN 60 241 GLU-RICH (ACIDIC).
SQ SEQUENCE 407 AA; 46617 MW; FFD399CA82CE136C CRC64;

Query Match 11.3%; Score 88; DB 1; Length 407;
Best Local Similarity 20.0%; Pred. No. 0.028;
Matches 25; Conservative 31; Mismatches 69; Indels 0; Gaps 0;
```

Qy	50	ATTAIEAASSDALEALADQTDLQSEEAUVVKADNAASDALEALADQTDALQSEEAEEVYQ
Dd	190	AESKSAATSGAGAKTSETSNASLSQASTATTTKASEAATSARDAAASKAAKSS
Qy	110	SDNAAAWAKEAATPIA 126
Dd	250	ETHNASSSASSASAATA 266
 RESULT 6 ID TIR3_YEAST STANDARD; PRT; 269 AA.		
AC	P40552;	
DT	01-FEB-1995 (Rel. 31, Created)	
DT	01-FEB-1995 (Rel. 31, Last sequence update)	
DT	01-MAR-2002 (Rel. 41, Last annotation update)	
DE	Cell wall protein TIR3 precursor.	
DE	TIR3 OR YILO1W.	
GN	Saccharomyces cerevisiae (Baker's yeast).	
OS	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.	
CX	NCBI_TaxId=4932;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=S286C / AB972;	
RA	Barrell B.G., Badcock K., Bankier A.T., Bowden S., Brown D.,	
RA	Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,	
RA	Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,	
RA	Louis E., Lyne G., Moule S., Moulton T., Odell C., Pearson D.,	
RA	Rajandream M.A., Riley L., Rowley N., Skelton J., Smith V.,	
RA	Walsh S.V., Whitehead S.;	
RL	Submitted (DEC-1994) to the EMBL/GenBank/DDJB databases.	
RP	[2]	
RN	REGULATION.	
RC	MEDLINE=21189315; PubMed=11292809;	
RA	Abranova N., Sertil O., Mehta S., Lowry C.V.;	
RT	"Reciprocal regulation of anaerobic and aerobic cell wall mannoprotein	
RL	gene expression in Saccharomyces cerevisiae.";	
RL	J. Bacteriol. 183:2881-2887(2001).	
CC	-1- FUNCTION: COMPONENT OF THE CELL WALL.	
CC	-1- PTM: EXTENSIVELY O-GLYCOSYLATED (By similarity).	
CC	-1- SIMILARITY: BELONGS TO THE SRP1 / TIP1 FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collabor-	
CC	ation between the Swiss Institute of Bioinformatics and the EMBL Outstat-	
CC	the European Bioinformatics Institute. There are no restrictions on	
CC	use by non-profit institutions as long as its content is in no	
CC	modified and this statement is not removed. Usage by and for comm-	
CC	entities requires a license agreement (See http://www.isb-sib.ch/anno-	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
EMBL	Z38113; CAAB6238.1; -	
DR	PIR; S48446; S48444.	
DR	SGD; S0001273; TIR3.	
DR	InterPro: IPR000992; SRP1_TIP1.	
DR	Pfam: PF00660; SRP1_TIP1.1.	
DR	PROSITE; PS00724; SRP1_TIP1; 1.	
KW	Cell wall; Glycoprotein; Signal.	
FT	STIGNAL 1 22 POTENTIAL.	
FT	CHAIN 23 269 CELL WALL PROTEIN TIR3.	
FT	DOMAIN 126 269 ALA/SER-RICH.	
FT	SEQUENCE 269 AA; 26308 MW; 48AF914D8C149124 CRC64;	
Qy	49	EATTAIEAASSDALEALADQTDLQSEEAUVVKADNAASDALEALADQTDALQSEEAEEVY 108
Dd	131	ESTATSSDASSGESSAASSSSASSSSASSSSASSSSASSSSASSSSASSSSASSSSA 190
Qy	109	QSDNAAADWAKEAATPIA 126

Db 191 KSSSAKSSGSSAASSNA 208

RESULT 7

TOLA\_PSEAE STANDARD; PRT; 347 AA.

ID TOLA\_PSEAE STANDARD; PRT; 347 AA.

AC P50600;

DT 01-OCT-1996 (Rel. 34, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE TOLA protein.

GN TOLA OR PA0971.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI\_TaxID=287;

RN [1]

RC SEQUENCE FROM N.A.

RP STRAIN-PAO;

RX MEDLINE-97113525; PubMed-8955385;

RT Dennis J.J., Lafontaine E.R., Sokol P.A.;

RT "Identification and characterization of the tolQRA genes of

RT Pseudomonas aeruginosa."

RL J. Bacteriol. 178:7059-7068(1996).

RN [2]

RP REVISIONS TO N-TERMINUS.

RA Duan K., Sokol P.A.;

RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RC SEQUENCE FROM N.A.

RP STRAIN-ATCC 15692 / PA01;

RX MEDLINE-20437337; PubMed-10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen."

RL Nature 406:959-964(2000).

RN [4]

RC FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS

CC (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane

CC (Potential).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: U39558; AAC44660.2; -.

DR EMBL; AE004530; AAG04360.1; -.

DR HSSP; P04002; 1WFA.

DR Transport; Protein transport; Transmembrane; Repeat; Inner membrane;

KW Complete proteome.

FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 17 37 POTENTIAL.

FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).

FT DOMAIN 209 216 POLY-ALA.

SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 10.3%; Score 80; DB 1; Length 347;

Best Local Similarity 29.8%; Pred. No. 0.18;

Matches 25; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

CC -----

CC EMBL: U39558; AAC44660.2; -.

DR EMBL; AE004530; AAG04360.1; -.

DR HSSP; P04002; 1WFA.

DR Transport; Protein transport; Transmembrane; Repeat; Inner membrane;

KW Complete proteome.

FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 17 37 POTENTIAL.

FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).

FT DOMAIN 209 216 POLY-ALA.

SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 10.3%; Score 80; DB 1; Length 347;

Best Local Similarity 29.8%; Pred. No. 0.18;

Matches 25; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

CC -----

CC EMBL: U39558; AAC44660.2; -.

DR EMBL; AE004530; AAG04360.1; -.

DR HSSP; P04002; 1WFA.

DR Transport; Protein transport; Transmembrane; Repeat; Inner membrane;

KW Complete proteome.

FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 17 37 POTENTIAL.

FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).

FT DOMAIN 209 216 POLY-ALA.

SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 10.3%; Score 80; DB 1; Length 347;

Best Local Similarity 29.8%; Pred. No. 0.18;

Matches 25; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

CC -----

CC EMBL: U39558; AAC44660.2; -.

DR EMBL; AE004530; AAG04360.1; -.

DR HSSP; P04002; 1WFA.

DR Transport; Protein transport; Transmembrane; Repeat; Inner membrane;

KW Complete proteome.

FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 17 37 POTENTIAL.

FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).

FT DOMAIN 209 216 POLY-ALA.

SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 10.3%; Score 80; DB 1; Length 347;

Best Local Similarity 29.8%; Pred. No. 0.18;

Matches 25; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

CC -----

CC EMBL: U39558; AAC44660.2; -.

DR EMBL; AE004530; AAG04360.1; -.

DR HSSP; P04002; 1WFA.

DR Transport; Protein transport; Transmembrane; Repeat; Inner membrane;

KW Complete proteome.

FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 17 37 POTENTIAL.

FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).

FT DOMAIN 209 216 POLY-ALA.

SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 10.3%; Score 80; DB 1; Length 347;

Best Local Similarity 29.8%; Pred. No. 0.18;

Matches 25; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

CC -----

CC EMBL: U39558; AAC44660.2; -.

DR EMBL; AE004530; AAG04360.1; -.

DR HSSP; P04002; 1WFA.

DR Transport; Protein transport; Transmembrane; Repeat; Inner membrane;

KW Complete proteome.

FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 17 37 POTENTIAL.

FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).

FT DOMAIN 209 216 POLY-ALA.

SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 10.3%; Score 80; DB 1; Length 347;

Best Local Similarity 29.8%; Pred. No. 0.18;

Matches 25; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

CC -----

CC EMBL: U39558; AAC44660.2; -.

DR EMBL; AE004530; AAG04360.1; -.

DR HSSP; P04002; 1WFA.

DR Transport; Protein transport; Transmembrane; Repeat; Inner membrane;

KW Complete proteome.

FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 17 37 POTENTIAL.

FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).

FT DOMAIN 209 216 POLY-ALA.

SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 10.3%; Score 80; DB 1; Length 347;

Best Local Similarity 29.8%; Pred. No. 0.18;

Matches 25; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

CC -----

CC EMBL: U39558; AAC44660.2; -.

DR EMBL; AE004530; AAG04360.1; -.

DR HSSP; P04002; 1WFA.

DR Transport; Protein transport; Transmembrane; Repeat; Inner membrane;

KW Complete proteome.

FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 17 37 POTENTIAL.

FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).

FT DOMAIN 209 216 POLY-ALA.

SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 10.3%; Score 80; DB 1; Length 347;

Best Local Similarity 29.8%; Pred. No. 0.18;

Matches 25; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

CC -----

CC EMBL: U39558; AAC44660.2; -.

DR EMBL; AE004530; AAG04360.1; -.

DR HSSP; P04002; 1WFA.

DR Transport; Protein transport; Transmembrane; Repeat; Inner membrane;

KW Complete proteome.

FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 17 37 POTENTIAL.

FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).

FT DOMAIN 209 216 POLY-ALA.

SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 10.3%; Score 80; DB 1; Length 347;

Best Local Similarity 29.8%; Pred. No. 0.18;

Matches 25; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

CC -----

CC EMBL: U39558; AAC44660.2; -.

DR EMBL; AE004530; AAG04360.

```

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL-WALL.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C40.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16421; CAA34442.1; ALT_INIT.
DR PIR: S05542; S05542.
DR MEROPS: C40.001; -.
DR InterPro: IPR000064; NLPC_P60.
DR Pfam: PF00877; NLPC_P60; 1.
KW Signal: Cell wall.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 516 P54 PROTEIN.
FT ACT_SITE 429 429 POTENTIAL.
SQ SEQUENCE 516 AA; 54596 MW; 402ECAA439846D26 CRC64;

Query Match 10.2%; Score 79; DB 1; Length 516;
Best Local Similarity 16.9%; Pred. No. 0.35; Indels 0; Gaps 0;
Matches 21; Conservative 30; Mismatches 73;

QY 26 EKALNRNEAIDELKKQAIQEDKATTAIFAASDSDALEALADQDALQSEAAVVKADNA 85
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 227 EQARIEQARLAEQARQAAQAEKAEKAEQAAQAAQATQALSSASITSESSAAQSSE 286
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
QY 86 ASDALEALADQDALQSEAAEVVSDNAASDAWEKATPTALDVKKTKPKVYKKEERQ 145
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 287 ESKAPESTTEESTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSS 346
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :

QY 146 NVNT 149
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 347 SSNT 350

RESULT 10
CALD_CHICK STANDARD; PRT; 771 AA.
ID P12957; Q90756; Q90761; Q92018; Q99230; Q03698;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Caldesmon (CDM).
GN CALD1 OR CAD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (GIZZARD H-CAD ISOFORM).
RC TISSUE=Gizzard, and Oviduct;
RX MEDLINE=89340480; PubMed=2760048;
RA Bryan J., Inai M., Lee R., Moore P., Cook R.G., Lin W.-G.;
RT "Cloning and expression of a smooth muscle caldesmon.";
RL J. Biol. Chem. 264:13873-13879(1989).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (GIZZARD H-CAD ISOFORM).
RC TISSUE=Gizzard;
RX MEDLINE=90026426; PubMed=2803315;
RA Hayashi K., Kanda K., Kimizuka F., Kato I., Sobue K.;
RT "Primary structure and functional expression of h-caldesmon
complementary DNA.";
RL Biochem. Biophys. Res. Commun. 164:503-511(1989).
RN [3]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (BRAIN L-CAD ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=91093148; PubMed=1824698;
RA Hayashi K., Fujio Y., Kato I., Sobue K.;

```

```

RT "Structural and functional relationships between h- and
l-caldesmons.";
RL J. Biol. Chem. 266:355-361(1991).
RN [4]
RP SEQUENCE FROM N.A. (GIZZARD H-CAD; BRAIN L-CAD AND GIZZARD L-CAD).
RX MEDLINE=94071934; PubMed=8250919;
RA Haruna M., Hayashi K., Yano H., Takeuchi O., Sobue K.;
RT "Common structural and expressional properties of vertebrate
caldesmon genes.";
RL Biochem. Biophys. Res. Commun. 197:145-153(1993).
RN [5]
RP SEQUENCE FROM N.A. (GIZZARD L-CAD ISOFORM).
RC TISSUE=Gizzard;
RX MEDLINE=92042686; PubMed=1939602;
RA Bryan J., Lee R.;
RT "Sequence of an avian non-muscle caldesmon.";
RL J. Muscle Res. Cell Motil. 12:372-375(1991).
RN [6]
RP SEQUENCE OF 1-15 FROM N.A. (BRAIN L-CAD ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=94271210; PubMed=8002994;
RA Yano H., Hayashi K., Haruna M., Sobue K.;
RT "Identification of two distinct promoters in the chicken caldesmon
gene.";
RL Biochem. Biophys. Res. Commun. 201:618-626(1994).
RN [7]
RP SEQUENCE OF 451-756 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Gizzard;
RX MEDLINE=89273666; PubMed=2730665;
RA Hayashi K., Yamada S., Kanda K., Kimizuka F., Kato I., Sobue K.;
RT "35 kDa fragment of h-caldesmon conserves two consensus sequences of
the tropomyosin-binding domain in troponin T.";
RL Biochem. Biophys. Res. Commun. 161:38-45(1989).
RN [8]
RP SEQUENCE OF 498-525.
RX MEDLINE=88293484; PubMed=3401222;
RA Mornet D., Audemard E., Derancourt J.;
RT "Identification of a 15 kilodalton actin binding region on gizzard
caldesmon probed by chemical cross-linking.";
RL Biochem. Biophys. Res. Commun. 154:564-571(1988).
RN [9]
RP PHOSPHORYLATION SITES.
RX MEDLINE=92041815; PubMed=1939059;
RA Mak A.S., Carpenter M., Smillie L.B., Wang J.H.;
RT "Phosphorylation of caldesmon by p34cdc2 kinase. Identification of
phosphorylation sites.";
RL J. Biol. Chem. 266:19971-19975(1991).
CC -1- FUNCTION: ACTIN- AND MYOSIN-BINDING PROTEIN IMPLICATED IN THE
REGULATION OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE AND
NONMUSCLE CELLS (COULD ACT AS A BRIDGE BETWEEN MYOSIN AND ACTIN
FILAMENTS). STIMULATES ACTIN BINDING OF TROPOMYOSIN WHICH
INCREASES THE STABILIZATION OF ACTIN FILAMENT STRUCTURE. IN MUSCLE
TISSUES, INHIBITS THE ACTOMYOSIN ATPASE BY BINDING TO P-ACTIN.
THIS INHIBITION IS ATTENUATED BY CALCIUM-CALMODULIN AND IS
POTENTIATED BY TROPOMYOSIN. INTERACTS WITH ACTIN, MYOSIN, TWO
MOLECULES OF TROPOMYOSIN AND WITH CALMODULIN. ALSO PLAY AN
ESSENTIAL ROLE DURING CELLULAR MITOSIS AND RECEPTOR CAPPING.
CC -1- SUBCELLULAR LOCATION: ON THIN FILAMENTS IN SMOOTH MUSCLE AND ON
STRESS FIBERS IN FIBROBLASTS (NONMUSCLE) (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; GIZZARD H-CAD (SHOWN HERE),
BRAIN L-CAD AND GIZZARD L-CAD; ARE PRODUCED BY ALTERNATIVE
SPLICING.
CC -1- TISSUE SPECIFICITY: HIGH-MOLECULAR-WEIGHT CALDESMON (H-CALDESMON)
IS PREDOMINANTLY EXPRESSED IN SMOOTH MUSCLES, WHEREAS LOW-
MOLECULAR-WEIGHT CALDESMON (L-CALDESMON) IS WIDELY DISTRIBUTED IN
NON-MUSCLE TISSUES AND CELLS. NOT EXPRESSED IN SKELETAL MUSCLE OR
HEART.
CC -1- DOMAIN: THE N-TERMINAL PART SEEMS TO BE A MYOSIN/CALMODULIN-
BINDING DOMAIN, AND THE C-TERMINAL A TROPOMYOSIN/ACTIN/CALMODULIN-
BINDING DOMAIN. THESE TWO DOMAINS ARE SEPARATED BY A CENTRAL
HELICAL REGION IN THE MUSCLE FORMS.
CC -1- PTM: IN NON-MUSCLE CELLS, PHOSPHORYLATION BY CDC2 DURING MITOSIS
CAUSES CALDESMON TO DISSOCIATE FROM MICROFILAMENTS.

```



Matches	22; Conservative	15; Mismatches	32; Indels	0; Gaps	0;				
QY	39	ELKQALIEDKATTAIPAASDALEALADQTDALQSEAAVAVKNAASDALEALADQTD	98						
Db	163	ERQDEIARQQAETAHLMAAAEAEKKNQDSILDEHRAVEAEKKAELAKAEAS	222						
QY	99	ALOSEEAEV	107						
Db	223	DVQSKQAIV	231						
RESULT 13									
XCPC_XENLA	STANDARD; PRT: 1290 AA.								
ID	XCPC_XENLA								
AC	P50532;								
DT	01-OCT-1996 (Rel. 34, Created)								
DT	01-OCT-1996 (Rel. 34, Last sequence update)								
DT	15-JUL-1998 (Rel. 36, Last annotation update)								
DE	Chromosome assembly protein XCAP-C.								
GN	XCAP-C.								
OS	Xenopus laevis (African clawed frog).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;								
OC	Xenopodidae; Xenopus.								
OX	NCBI_TaxID=8355;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=95042742; PubMed=7954811;								
RT	Hirano T., Mitchison T.J.;								
RA	"A heterodimeric coiled-coil protein required for mitotic chromosome								
RL	condensation in vitro."								
RL	Cell 79:449-458(1994).								
CC	-1- FUNCTION: REQUIRED FOR BOTH ASSEMBLY AND STRUCTURAL MAINTENANCE OF								
CC	MITOTIC CHROMOSOMES.								
CC	-1- SUBUNIT: ASSOCIATES WITH XCAP-E PROBABLY AS HETERODIMER.								
CC	-1- SUBCELLULAR LOCATION: NUCLEAR. DURING CHROMOSOME ASSEMBLY IN								
CC	MITOTIC EXTRACTS, XCAP-C/E WAS RECRUITED TO THE CHROMATIN AND								
CC	FORMED A DISCRETE INTERNAL STRUCTURE WITHIN ASSEMBLED CHROMOSOMES.								
CC	-1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS								
CC	PLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.								
CC	-1- SIMILARITY: BELONGS TO THE SMC FAMILY.								
CC	-----								
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -								
CC	the European Bioinformatics Institute. There are no restrictions on its								
CC	use by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>								
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).								
CC	-----								
DR	EMBL: U13673; AAA64679.1; "								
DR	InterPro: IPR003439; ABC_transportr.								
DR	InterPro: IPR003405; SMC_C.								
DR	InterPro: IPR003395; SMC_N.								
DR	Pfam: PF02483; SMC_C; 1.								
DR	Pfam: PF02453; SMC_N; 1.								
KW	Mitosis; ATP-binding; Coiled								
FT	NE_BIND 107 114								
FT	DOMAIN 264 594								
FT	COILED COIL (POTENTIAL).								
FT	DOMAIN 764 1027								
FT	COILED COIL (POTENTIAL).								
FT	DOMAIN 1094 1129								
FT	COILED COIL (POTENTIAL).								
FT	DOMAIN 1263 1290								
FT	COILED COIL (POTENTIAL).								







CC TRANSPORT OF VESICLES AND ORGANELLES.  
CC P150 (SUBUNIT) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC  
CC DYNEIN.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; P150 (SHOWN HERE) AND P135;  
CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: BRAIN.  
CC -!- PTM: PHOSPHORYLATED.  
CC -!- SIMILARITY: BELONGS THE DYNACTIN 150 KDA SUBUNIT FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF064205; AAD55811.1; JOINED.  
DR EMBL; AF064203; AAD55811.1; JOINED.  
DR EMBL; AF064204; AAD55811.1; JOINED.  
DR EMBL; X98801; CAA67333.1; JOINED.  
DR EMBL; AF066947; AAD03694.1; JOINED.  
DR EMBL; AF066927; AAD03694.1; JOINED.  
DR EMBL; AF066928; AAD03694.1; JOINED.  
DR EMBL; AF066929; AAD03694.1; JOINED.  
DR EMBL; AF066930; AAD03694.1; JOINED.  
DR EMBL; AF066931; AAD03694.1; JOINED.  
DR EMBL; AF066932; AAD03694.1; JOINED.  
DR EMBL; AF066933; AAD03694.1; JOINED.  
DR EMBL; AF066934; AAD03694.1; JOINED.  
DR EMBL; AF066935; AAD03694.1; JOINED.  
DR EMBL; AF066936; AAD03694.1; JOINED.  
DR EMBL; AF066937; AAD03694.1; JOINED.  
DR EMBL; AF066938; AAD03694.1; JOINED.  
DR EMBL; AF066939; AAD03694.1; JOINED.  
DR EMBL; AF066940; AAD03694.1; JOINED.  
DR EMBL; AF066941; AAD03694.1; JOINED.  
DR EMBL; AF066942; AAD03694.1; JOINED.  
DR EMBL; AF066943; AAD03694.1; JOINED.  
DR EMBL; AF066944; AAD03694.1; JOINED.  
DR EMBL; AF066945; AAD03694.1; JOINED.  
DR EMBL; AF066946; AAD03694.1; JOINED.  
DR MIM; 601143; -;  
DR InterPro; IPR000938; CAP-Gly.  
DR Pfam; PF01302; CAP\_GLY; 1.  
DR PROSITE; PS00845; CAP\_GLY\_1; 1.  
DR PROSITE; PS0245; CAP\_GLY\_2; 1.  
KW Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton;  
KW Alternative splicing; Phosphorylation.  
FT DOMAIN 48 90  
FT CAP-Gly.  
FT SER-RICH.  
FT COILED COIL (POTENTIAL).  
FT COILED COIL (POTENTIAL).  
FT COILED COIL (POTENTIAL).  
FT COILED COIL (POTENTIAL).  
FT MAOSKRHYVSRTPSGSRMSAEASARPLRVGSRVVEVIGKGRH  
FT GIVAVIGATLFAIGKVGVLDEAGKNDGTGVQGRKRYFTCD  
FT EHGIFVRSQIQVFDGADFTSPETPDSSASKVLKREGTD  
FT TTKATSKSL -> MMRQ (IN ISOFORM P135).  
FT A -> P.  
FT /FTID=VAR\_001373.  
FT S -> N (IN REF. 2 AND 3).  
FT MISSING (IN REF. 2 AND 3).  
FT D -> V (IN REF. 2 AND 3).  
FT VARSPLIC 1 131  
FT VARIANT 163 163  
FT CONFLICT 10 10  
FT CONFLICT 132 138  
FT CONFLICT 712 712  
FT CONFLICT 712 712  
FT SEQUENCE 1278 AA; 141694 MW; 6DCEA5E67856E4BC CRC64;  
Query Match 9.4%; Score 73; DB 1; Length 1278;  
Best Local Similarity 24.4%; Pred. No. 4.1;  
Matches 19; Conservative 20; Mismatches 39; Indels 0; Gaps 0;  
QY 38 DELKKQAIKEDKATTATIEAASSDALEALADQTDALQSEAAVVKADNAASDALEALADQTD 97

Db 964 EELSEANVRLSLLEKLDAAKADADERIEKVQTRLEETQALLRKKEFEETMDALQADI 1023  
QY 98 DALQSEAEVQSDNAAS 115  
Db 1024 DQLEAKAEKQRLNASQS 1041  
RESULT 18  
VLTF\_BPT5 STANDARD; PRT; 1396 AA.  
AC P13390; O48502;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-SEP-1996 (Rel. 33, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE L-shaped tail fiber protein (LTF protein).  
GN LTF.  
OS Bacteriophage T5.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
OX NCBI\_TaxID=10726;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95309401; PubMed=7789514;  
RA Kallman A.V., Kulshin V.E., Shiyapnikov M.G., Ksenzenko V.N.,  
RA Kryukov V.M.;  
RT "The nucleotide sequence of the bacteriophage T5 ltf gene.";  
RL FEBS Lett. 366:46-48(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kallman A.V.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE=88289370; PubMed=3267228;  
RA Kallman A.V., Kryukov V.M., Bayev A.A.;  
RT "The nucleotide sequence of bacteriophage T5 DNA at the region  
RT between early and late genes.";  
RL Nucleic Acids Res. 16:6230-6230(1988).  
CC -!- FUNCTION: NONESSENTIAL PROTEIN THAT MEDIATES BINDING TO THE  
CC POLYMANNOSE O ANTIGEN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X69460; CAA49220.1; -;  
DR EMBL; AJ001191; CAA04591.1; -;  
DR PIR; S01982; S01982.  
KW Late protein.  
FT CONFLICT 986 986 V -> A (IN REF. 2).  
FT SEQUENCE 1396 AA; 147989 MW; 18CD2192F65FFFC1 CRC64;  
Query Match 9.4%; Score 73; DB 1; Length 1396;  
Best Local Similarity 29.5%; Pred. No. 4.5;  
Matches 23; Conservative 13; Mismatches 42; Indels 0; Gaps 0;  
QY 49 EATTATEAASSDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAEV 108  
Db 91 QSASSATANSASAKAATSETNANNKNAKTSETNAASSASSASSFATAENSARA 150  
QY 109 QSDNAASDAWEKAATPIA 126  
Db 151 SETNAGNSAQADADAKSTA 168  
RESULT 19  
AK15\_RAT STANDARD; PRT; 464 AA.  
ID AK15\_RAT  
AC P24587;





RN SEQUENCE FROM N.A.  
 RP MEDLINE-94183253; PubMed-8135819;  
 RA Kitajima Y., Yatsuki H., Zhang R., Matsubashi S., Hori K.;  
 RT "A novel human homologue of a dead-box RNA helicase family.";  
 RL Biochem. Biophys. Res. Commun. 199:748-754(1994).  
 CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: D26528; BAA05534.1; -  
 DR MIM: 603793; -  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR000629; DEAD\_ATP\_helicase.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF00270; DEAD; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00490; HELIC; 1.  
 DR PROSITE: PS00039; DEAD\_ATP\_HELICASE; FALSE\_NEG.  
 KW Hydrolyase; ATP-binding; Helicase; RNA-binding.  
 FT NP\_BIND 39 46 ATP (POTENTIAL).  
 FT SITE 145 148 DEAD BOX  
 SQ SEQUENCE 439 AA; 47405 MW; EA9E81631D5D0847 CRC64;  
  
 Query Match 9.1%; Score 71; DB 1; Length 439;  
 Best Local Similarity 26.8%; Pred. No. 2.3;  
 Matches 19; Conservative 14; Mismatches 38; Indels 0; Gaps 0;  
  
 QY 75 EEAAYKADNAASDALEALADQTDALQSEAEVQSDNAASDAWEKAATPIALDVKTKD 134  
 DB 349 DDFVLVKAATFADAKQATPKSTAAVKAARATAPTDQAGHQAGQAQKPTQRTTQAK 408  
  
 QY 135 TKPVVKKEERQ 145  
 DB 409 TVPVVKPKKK 419  
  
 RESULT 24  
 SBCC\_RHOCA STANDARD; PRT; 1238 AA.  
 ID SBCC\_RHOCA  
 AC Q56012;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DE Transcription initiation factor TFIID subunit C.  
 GN SBCC.  
 OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Rhodobacter.  
 OX NCBI\_TaxID=1061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SB1003 / St Louis;  
 RX MEDLINE-97040404; PubMed-9256491;  
 RA Vicek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;  
 RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter  
 capsulatus SB1003.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).  
 CC -1- FUNCTION: SbcC cleaves DNA hairpin structures. These structures  
 CC can inhibit DNA replication and are intermediates in certain DNA  
 CC recombination reactions. The complex acts as a 3'-5' double  
 CC strand exonuclease that can open hairpins. It also has a 5'  
 CC single-strand endonuclease activity (By similarity).  
 CC -1- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: AF010496; AAC16118.1; -  
 DR InterPro: IPR003439; ABC\_transport.  
 DR InterPro: IPR001687; ATP\_GTP\_A.  
 KW Hydrolyase; Nuclease; Exonuclease; Endonuclease; DNA replication;  
 KW DNA recombination; ATP-binding; Coiled coil.  
 FT NP\_BIND 37 44 ATP (POTENTIAL).  
 FT DOMAIN 395 438 COILED COIL (POTENTIAL).  
 FT DOMAIN 466 487 COILED COIL (POTENTIAL).  
 FT DOMAIN 521 600 COILED COIL (POTENTIAL).  
 FT DOMAIN 724 770 COILED COIL (POTENTIAL).  
 FT DOMAIN 901 943 COILED COIL (POTENTIAL).  
 FT DOMAIN 1019 1052 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 1238 AA; 128046 MW; 2B6BC6C63A859AE1 CRC64;  
  
 Query Match 9.1%; Score 71; DB 1; Length 1238;  
 Best Local Similarity 33.3%; Pred. No. 6.7;  
 Matches 20; Conservative 9; Mismatches 31; Indels 0; Gaps 0;  
  
 QY 42 KOAIEDKEATTATGAASSDALEALADQTDALQSEAAVAVKADNAASDALEALADQTDALQ 101  
 DB 447 RRLADHRAACAQAQAATAAQAQDRAQAQAQAAAAAQAQAAAAAQAQAAAAAQAQ 506  
  
 RESULT 25  
 T2D9\_HUMAN STANDARD; PRT; 211 AA.  
 ID T2D9\_HUMAN  
 AC Q15344;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DE Transcription initiation factor TFIID subunit C.  
 GN TAFII28 (TFIID subunit p30-beta).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE-95246745; PubMed-7729427;  
 RA Mengus G., May M., Jacq X., Staub A., Tora L., Chambon P.,  
 RA Davidson I.;  
 RT "Cloning and characterization of hTAFII18, hTAFII20 and hTAFII28:  
 three subunits of the human transcription factor TFIID.";  
 RL EMBO J. 14:1520-1531(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-96418138; PubMed-8820923;  
 RA Kuzubara T., Horikoshi M.;  
 RT "Isolation and characterization of a cDNA encoding a human TFIID  
 subunit containing a variety of putative structural motifs including  
 direct repeats.";  
 RL Biol. Pharm. Bull. 19:122-126(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Williams S.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CORE TAFII PRESENT IN BOTH OF THE PREVIOUSLY DESCRIBED  
 CC TFIID SPECIES WHICH EITHER LACK OR CONTAIN TAFII30 (TFIID ALPHA  
 CC AND TFIID BETA RESPECTIVELY).  
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A  
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). INTERACTS WITH TAFII18  
 CC BOTH IN VITRO AND INTRACELLULARLY; ALSO INTERACTS DIRECTLY WITH  
 CC TBP.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -!- DOMAIN: TBP AND TAFII18 BIND TO DISTINCT DOMAINS OF TAFII28.  
CC -!- SIMILARITY: BELONGS TO THE TAF2I FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X83928; CAA58780.1; -  
CC EMBL: D63705; BAA23620.1; -  
CC EMBL: AL033520; CAB96725.1; -  
CC TRANSFAC: T02114; -  
CC MIM: 600772; -  
CC InterPro: IPR000166; Histone\_core.  
KW Transcription regulation; Nuclear protein.  
SQ SEQUENCE 211 AA; 23307 MW; 77C7BDC667D19A8F CRC64;  
  
Query Match 9.0%; Score 70; DB 1; Length 211;  
Best Local Similarity 21.6%; Pred. No. 1.4;  
Matches 22; Conservative 21; Mismatches 59; Indels 0; Gaps 0;  
  
QY 55 EAASSDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVYQSDNAA 114  
DB 14 ETGESDetaVPCDPCATDGTGDIPEETDGDADVDLKEAAAECELESDQVSDLTIVERED 73  
  
QY 115 SDAWKAATPIALDVKKTKDTPVKKKEERQNVTLPTTGE 156  
DB 74 SSLNPAKKLKIDTKRKKKQKVDEDETKMQILVSSFSE 115  
  
RESULT 26  
TUB8\_SOLUTU STANDARD; PRT; 211 AA.  
AC P33191;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE Induced stolon tip protein TUB8 (Fragment).  
GN TUB8.  
OS Solanum tuberosum (Potato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4113;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=CV. RECORD; TISSUE=Stolon tip;  
RX MEDLINE=93081725; PubMed=1450379;  
RA Taylor M.A., Mad Arif S.A., Kumar A., Davies H.V., Scobie L.A.,  
RA Pearce S.R., Flavell A.J.;  
RT "Expression and sequence analysis of cDNAs induced during the early  
RT stages of tuberisation in different organs of the potato plant  
RT (Solanum tuberosum L.).";  
RL Plant Mol. Biol. 20:641-651(1992).  
CC -!- TISSUE SPECIFICITY: STOLON, ALSO EXPRESSED IN LEAVES, STEMS  
CC AND ROOTS.  
CC  
CC -!- DEVELOPMENTAL STAGE: TRANSCRIBED IN THE STOLON TIP DURING THE  
CC EARLY STAGES OF TUBERIZATION. MAXIMUM EXPRESSION WAS IN NON-  
CC SWELLING STOLON TIPS FROM STAGE B, AND LEVEL REMAINED HIGH AS  
CC TUBER INCREASED IN SIZE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: 211679; CAA77741.1; -

DR PIR; S28046; S28046.  
KW Repeat.  
FT NON\_TER 1  
FT DOMAIN 56 141 9 X 6-7 AA REPEATS OF E-E-P-A-A-A.  
FT REPEAT 56 61 1 (APPROXIMATE).  
FT REPEAT 76 81 2 (APPROXIMATE).  
FT REPEAT 84 88 3 (APPROXIMATE).  
FT REPEAT 92 97 4.  
FT REPEAT 107 112 5 (APPROXIMATE).  
FT REPEAT 115 120 6.  
FT REPEAT 121 126 7.  
FT REPEAT 127 133 8 (APPROXIMATE).  
FT REPEAT 136 140 9 (APPROXIMATE).  
FT DOMAIN 203 208 POLY-TYR.  
SQ SEQUENCE 211 AA; 22638 MW; AFF293P819951C4B CRC64;  
  
Query Match 9.0%; Score 70; DB 1; Length 211;  
Best Local Similarity 23.2%; Pred. No. 1.4;  
Matches 26; Conservative 19; Mismatches 67; Indels 0; Gaps 0;  
  
QY 45 IEDKEATTAEAASSDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEE 104  
DB 62 VEKEVEVESAPVVEEAEAPVAAEAAALVAEPAEAAEPTAAVAANVEPVAAPVEEPAAE 121  
  
QY 105 AEYVQSDNAAADAWKAATPIALDVKKTKDTPVKKKEERQNVTLPTTGE 156  
DB 122 EPAAEPEPVAAPVVEEAEAPVAAEAPVSEPEAEKAEASPVSEEPKVEE 173  
  
RESULT 27  
TOLA\_ECOLI STANDARD; PRT; 421 AA.  
AC P19934;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tola protein.  
GN TOLA OR CIM OR EXCC OR LKY OR B0739.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=JM105;  
RX MEDLINE=90078104; PubMed=2687247;  
RA Levensgood S.K., Webster R.E.;  
RT "Nucleotide sequences of the tola and tolB genes and localization of  
RT their products, components of a multistep translocation system in  
RT Escherichia coli.";  
RL J. Bacteriol. 171:6600-6609(1989).  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
[3]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 12.7-28.0 min region on the linkage map.";







```

RA Zopf D., Dineva B., Betz H., Gundelfinger E.D.;
RT "Isolation of the chicken middle-molecular weight neurofilament
RL (NF-M) gene and characterization of its promoter.";
RN Nucleic Acids Res. 18:521-529(1990).
RP [2]
RX SEQUENCE OF 259-857 FROM N.A.
RA MEDLINE=88112814; PubMed=3123320;
RA Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D., Betz H.;
RT "Identification of gene products expressed in the developing chick
RT visual system: characterization of a middle-molecular-weight
RL neurofilament cDNA.";
RL Genes Dev. 1:699-708(1987).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NEM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17102; CAA34958.1; -
DR EMBL; X05558; CAA29073.1; -
DR PIR; A27040; A27040.
DR PIR; S08061; S08061.
DR PIR; S15762; S15762.
DR InterPro: IPR001664; IF.
DR Pfam: PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW Glycoprotein.
FT INIT_MET 0 0
FT DOMAIN 1 98 HEAD.
FT DOMAIN 99 406 ROD.
FT DOMAIN 407 857 TAIL.
FT DOMAIN 99 130 COIL 1A.
FT DOMAIN 131 143 LINKER 1.
FT DOMAIN 144 242 COIL 1B.
FT DOMAIN 243 259 LINKER 12.
FT DOMAIN 260 281 COIL 2A.
FT DOMAIN 282 285 LINKER 2.
FT DOMAIN 286 406 COIL 2B.
FT CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 426 426 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CONFLICT 546 546 G -> R (IN REF. 2).
SQ SEQUENCE 857 AA; 95704 MW; 4E2E0FC6AC64778B CRC64;

Query Match 9.0%; Score 70; DB 1; Length 857;
Best Local Similarity 23.4%; Pred. No. 5.9;
Matches 22; Conservative 23; Mismatches 49; Indels 0; Gaps 0;

QY 30 LRNEERAIDELKKQAEIDKKAIEDKATTAIEAASSDALEALADQTDALQSEEAAYKADNAASDA 89
Db 442 LKQVHKFVEEIEETKVEESEMEDALSATAEAAKQAEQEEKEAEAEAEVEEAVS 501
QY 90 LEALADQTDALQSEEAAYVQSDNAASDAWEKAAT 123
Db 502 EKAAQAAEEKEEKEEAEAEAEAEKSDAEEGS 535

RESULT 30

```

```

MST2_DROHY
ID MST2_DROHY STANDARD; PRT; 1391 AA.
AC Q08696;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein mst101(2).
GN MST101(2).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=95045538; PubMed=7957199;
RA Neesen J., Padmanabhan S., Buenemann H.;
RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid
RT motif representing the major component of the sperm-tail-specific
RT axoneme-associated protein family Dhmst101 form extended
RT alpha-helical rods within the extremely elongated spermatozoa of
RT Drosophila hydei.";
RL Eur. J. Biochem. 225:1089-1095(1994).
CC -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
CC SPERMATIDS.
CC -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -!- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
CC REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73481; CAA51876.1; -
DR PIR; S34154; S34154.
DR HSSP; P01032; IG5A.
DR FlyBase: FBgn020733; Dhyd\mst101(2).
KW Sperm; Repeat; Multigene family; Polymorphism.
FT DOMAIN 332 1268 [KR]-K-X-C-X-X-K-X-K-X-K-X-X-X-E.
FT SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48878 CRC64;

Query Match 9.0%; Score 70; DB 1; Length 1391;
Best Local Similarity 28.2%; Pred. No. 9.7;
Matches 20; Conservative 13; Mismatches 38; Indels 0; Gaps 0;

QY 22 GNAPEKALRNDEERAIDELKKQAEIDKKAIEDKATTAIEAASSDALEALADQTDALQSEEAAYVK 81
Db 753 GNGKGGKALKKEKKRELAKKAAEKKEKKAEEKKEKKAEEKKEKKEKTAKKRKEAEKKK 812
QY 82 ADNAASDALEA 92
Db 813 CEKTAKKRKEA 823

RESULT 31
HBHA_MYCTU
ID HBHA_MYCTU STANDARD; PRT; 198 AA.
AC Q11142; O85733;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heparin-binding hemagglutinin (Adhesin).
GN HBHA OR RV0475 OR MT0493 OR MTCY20G9.01.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.

```



**THIS PAGE BLANK (USPTO)**

; LENGTH: 16135  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-09-614-150-35057

Alignment Scores:  
Pred. No.: 2.29e+03 Length: 16135  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-614-150-35057 (1-16135)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
Db 5971 CTGGAGGCTCTTCAAGCAGGTGGAGTCTCTCAAGCCGCGAGTTGCTCTCTTGTGGCA 5912  
Qy 24 AlaAspAsn 26  
Db 5911 GCCGACAA 5903

## RESULT 40

US-09-619-049-1340/c  
; Sequence 1340, Application US/09619049  
; GENERAL INFORMATION:  
; APPLICANT: YANDELL, MARK  
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL  
; TITLE OF INVENTION: FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL  
; TITLE OF INVENTION: DROSOPHILA PROTEINS, AND USES THEREOF AS INSECTICIDAL  
; TITLE OF INVENTION: TARGETS  
; FILE REFERENCE: CL000735  
; CURRENT APPLICATION NUMBER: US/09/619,049  
; CURRENT FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: 60/171,590  
; PRIOR FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: 60/171,627  
; PRIOR FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: 60/175,763  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/175,685  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/186,663  
; PRIOR FILING DATE: 2000-03-03  
; PRIOR APPLICATION NUMBER: 60/187,241  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 1533  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1340  
; LENGTH: 16135  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-09-619-049-1340

Alignment Scores:  
Pred. No.: 2.29e+03 Length: 16135  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-619-049-1340 (1-16135)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
Db 5971 CTGGAGGCTCTTCAAGCAGGTGGAGTCTCTCAAGCCGCGAGTTGCTCTCTTGTGGCA 5912  
Qy 24 AlaAspAsn 26  
Db 5911 GCCGACAA 5903

Search completed: October 13, 2002, 04:38:41  
Job time : 564.612 secs

ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1230)  
OTHER INFORMATION: any n = a, g, c, t, unknown, or other  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1230)  
OTHER INFORMATION: 40182 : FINISHED (Clone Number : FINISHED)  
US-09-565-309A-62025

Alignment Scores:  
Pred. No.: 116 Length: 1230  
Score: 52.00 Matches: 11  
Percent Similarity: 72.73% Conservatives: 5  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 40.31% Indels: 0  
DB: 22 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-565-309A-62025 (1-1230)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
||| |||||:||||: ||| ||| :|||:||||: |||  
Db 518 CTTCGGCAGCTCTCGGACAAAGGACATGCTCAATTCAAAAGTGCATCGCAAG 459

Qy 24 AlaAsp 25  
|||||  
Db 458 GCGGAT 453

RESULT 37

US-09-595-329A-545/c  
Sequence 545, Application US/09595329A

GENERAL INFORMATION:

APPLICANT: ALEXANDROV, Nikolai

APPLICANT: VYACHESLAV, Brover

TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide

TITLE OF INVENTION: Thereby

FILE REFERENCE: 2750-0948P

CURRENT APPLICATION NUMBER: US/09/595,329A

CURRENT FILING DATE: 2000-06-16

NUMBER OF SEQ ID NOS: 3309

SOFTWARE: PatentIn version 3.0

SEQ ID NO 545

LENGTH: 1230

TYPE: DNA

ORGANISM: Arabidopsis Thaliana

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (1)..(1230)

OTHER INFORMATION: Ceres Seq. ID no. 1013068

NAME/KEY: misc\_feature

LOCATION: (1)..(1230)

OTHER INFORMATION: n is a, c, t, g, unknown, or other

US-09-595-329A-545

Alignment Scores:  
Pred. No.: 116 Length: 1230  
Score: 52.00 Matches: 11  
Percent Similarity: 72.73% Conservatives: 5  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 40.31% Indels: 0  
DB: 22 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-595-329A-545 (1-1230)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
||| |||||:||||: ||| ||| :|||:||||: |||  
Db 518 CTTCGGCAGCTCTCGGACAAAGGACATGCTCAATTCAAAAGTGCATCGCAAG 459

Qy 24 AlaAsp 25  
|||||  
Db 458 GCGGAT 453

RESULT 38

US-09-565-309A-52656/c  
Sequence 52656, Application US/09565309A  
GENERAL INFORMATION:  
APPLICANT: ALEXANDROV, Nikolai  
APPLICANT: BROVER, Vyacheslav  
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
TITLE OF INVENTION: THEREBY  
FILE REFERENCE: 2750-0853P  
CURRENT APPLICATION NUMBER: US/09/565,309A  
CURRENT FILING DATE: 2000-05-05  
NUMBER OF SEQ ID NOS: 68449  
SEQ ID NO 52656  
LENGTH: 1244  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1244)  
OTHER INFORMATION: any n = a, g, c, t, unknown, or other  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1244)  
OTHER INFORMATION: 40182 : 5TAG CONSENSUS (Clone Number:5tag\_consensus)  
US-09-565-309A-52656

Alignment Scores:  
Pred. No.: 117 Length: 1244  
Score: 52.00 Matches: 11  
Percent Similarity: 72.73% Conservatives: 5  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 40.31% Indels: 0  
DB: 22 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-565-309A-52656 (1-1244)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
||| |||||:||||: ||| ||| :|||:||||: |||  
Db 532 CTTCGGCAGCTCTCGGACAAAGGACATGCTCAATTCAAAAGTGCATCGCAAG 473

Qy 24 AlaAsp 25  
|||||  
Db 472 GCGGAT 467

RESULT 39

US-09-614-150-35057/c

Sequence 35057, Application US/09614150

GENERAL INFORMATION:

APPLICANT: Venter, J. Craig

APPLICANT: et al.

TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID

TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE

TITLE OF INVENTION: DROSOPHILA GENES.

FILE REFERENCE: CL000728

CURRENT APPLICATION NUMBER: US/09/614,150

CURRENT FILING DATE: 2000-07-11

PRIOR APPLICATION NUMBER: 60/157,832

PRIOR FILING DATE: 1999-10-05

PRIOR APPLICATION NUMBER: 60/160,191

PRIOR FILING DATE: 1999-10-19

PRIOR APPLICATION NUMBER: 60/161,932

PRIOR FILING DATE: 1999-10-28

PRIOR APPLICATION NUMBER: 60/164,769

PRIOR FILING DATE: 1999-11-12

PRIOR APPLICATION NUMBER: 60/173,383

PRIOR FILING DATE: 1999-12-28

PRIOR APPLICATION NUMBER: 60/175,693

PRIOR FILING DATE: 2000-01-12

PRIOR APPLICATION NUMBER: 60/184,831

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/191,637

PRIOR FILING DATE: 2000-03-23

NUMBER OF SEQ ID NOS: 43008

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 35057

```
Db 401 GCGGAT 396
RESULT 33
US-09-565-309A-67148/c
; Sequence 67148, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-0853P
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 67148
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(939)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc.feature
; LOCATION: (1)..(939)
; OTHER INFORMATION: 40182 : CDS (Clone Number : CDS)
US-09-565-309A-67148
Alignment Scores:
Pred. No.: 84.7 Length: 939
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-67148 (1-939)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 382 CTTCCGCGCACTCTCCGACAAAGGACATGCCTCAAAATTCAAAAGCTGCAATCGCAAG 323

QY 24 AlaAsp 25
Db 322 GCGGAT 317

RESULT 34
US-09-708-427-29368/c
; Sequence 29368, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29368
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..984
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc.feature
; LOCATION: 1..984
; OTHER INFORMATION: Ceres Seq. ID 1825505
US-09-708-427-29368
Alignment Scores:
Pred. No.: 89.5 Length: 984
Score: 52.00 Matches: 11
```

```
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 28 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-708-427-29368 (1-984)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 442 CTTCCGCGCACTCTCCGACAAAGGACATGCCTCAAAATTCAAAAGCTGCAATCGCAAG 383

QY 24 AlaAsp 25
Db 382 GCGGAT 377

RESULT 35
US-09-565-309A-67147/c
; Sequence 67147, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-0853P
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 67147
; LENGTH: 999
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(999)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc.feature
; LOCATION: (1)..(999)
; OTHER INFORMATION: 40182 : CDS (Clone Number : CDS)
US-09-565-309A-67147
Alignment Scores:
Pred. No.: 91 Length: 999
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-67147 (1-999)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 442 CTTCCGCGCACTCTCCGACAAAGGACATGCCTCAAAATTCAAAAGCTGCAATCGCAAG 383

QY 24 AlaAsp 25
Db 382 GCGGAT 377

RESULT 36
US-09-565-309A-62025/c
; Sequence 62025, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-0853P
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 62025
; LENGTH: 1230
; TYPE: DNA
```

```
* ; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17589
; LENGTH: 7206
; TYPE: DNA
; ORGANISM: Drosophila
US-60-167-217-17589

Alignment Scores:
Pred. No.: 402 Length: 7206
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 55 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-167-217-17589 (1-7206)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
Db 4905 CAGGAGGAGCTGACGAGGAGGAGCGCGCGGAGGAGGATGCGCGCGGCTGCGCGC 4964
Qy 25 AspAsnAlaAla 28
Db 4965 GATGCTGCGCGC 4976

RESULT 30
US-09-528-237A-1811/c
; Sequence 1811, Application US/09528237A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: The Drosophila Genome, Primary Nucleic
; TITLE OF INVENTION: Acid Sequences, Systems Containing The Nucleic Acid
; TITLE OF INVENTION: Sequences and Uses Thereof
; FILE REFERENCE: C1000284
; CURRENT APPLICATION NUMBER: US/09/528.237A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 2926
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1811
; LENGTH: 32275
; TYPE: DNA
; ORGANISM: Drosophila
US-09-528-237A-1811

Alignment Scores:
Pred. No.: 2,29e+03 Length: 32275
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 19 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-528-237A-1811 (1-32275)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
Db 26995 GAGGAGGAGCTGACGAGGAGGAGCGCGCGGAGGAGGATGCGCGCGGCTGCGCGC 26936
Qy 25 AspAsnAlaAla 28
Db 26935 GATGCTGCGCGC 26924

RESULT 31
US-09-565-309A-32821/c
; Sequence 32821, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-32821 (1-594)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLys 23
Db 518 CTTCCGCGACTCTCCGACAAAGGAGGACATGCTCAAAATTCAAAAGCTGCAATCGCAAG 459
Qy 24 AlaAsp 25
Db 458 GCGGAT 453

RESULT 32
US-09-565-309A-32824/c
; Sequence 32824, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 32824
; LENGTH: 641
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(641)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(641)
; OTHER INFORMATION: 40182:957281 (Clone Number:Unique Sequence Identifier)
US-09-565-309A-32824

Alignment Scores:
Pred. No.: 54.4 Length: 641
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-32824 (1-641)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLys 23
Db 461 CTTCCGCGACTCTCCGACAAAGGAGGACATGCTCAAAATTCAAAAGCTGCAATCGCAAG 402
Qy 24 AlaAsp 25
Db 461 CTTCCGCGACTCTCCGACAAAGGAGGACATGCTCAAAATTCAAAAGCTGCAATCGCAAG 402
```

```
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-09-614-150-17530

Alignment Scores:
Pred. No.: 350      Length: 6398
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 23      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-614-150-17530 (1-6398)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: |||||
Db 2298 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGAGGATGCCGAGCGGCTGCCGCC 2239

QY 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTCGCGCG 2227

RESULT 26
US-60-191-637-17580/c
; Sequence 17580, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; FILE REFERENCE: CL000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17580
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-637-17580

Alignment Scores:
Pred. No.: 350      Length: 6398
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 58      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-191-637-17580 (1-6398)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: |||||
Db 2298 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGAGGATGCCGAGCGGCTGCCGCC 2239

QY 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTCGCGCG 2227

RESULT 27
US-60-191-681-13896/c
; Sequence 13896, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: Li, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; FILE REFERENCE: c1000390
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13896
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-681-13896

Alignment Scores:
Pred. No.: 350      Length: 6398
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 58      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-191-681-13896 (1-6398)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: |||||
Db 2298 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGAGGATGCCGAGCGGCTGCCGCC 2239

QY 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTCGCGCG 2227

RESULT 28
US-60-173-464-14400
; Sequence 14400, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14400
; LENGTH: 7203
; TYPE: DNA
; ORGANISM: Drosophila
US-60-173-464-14400

Alignment Scores:
Pred. No.: 402      Length: 7203
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 56      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-173-464-14400 (1-7203)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: |||||
Db 4905 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGAGGATGCCGAGCGGCTGCCGCC 4964

QY 25 AspAsnAlaAla 28
   ||| |||||
Db 4965 GATGCTCGCGCG 4976

RESULT 29
US-60-167-217-17589
; Sequence 17589, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
```



**Qy** 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValLysAla 24  
||| ||||||:: ||||| :::::||||| |||  
**Db** 1122 GAGCAGGAAGCTCACGAGGAGGACGCCGGCAGGCAGAGATGCCAGCGCTGCCCC 1063

[illegible]

```
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Drosophila
US-60-142-845-289

Alignment Scores:
Pred. No.: 79.6 Length: 1783
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 53 Caps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-142-845-289 (1-1783)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1475 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGACAGGATGCCGCGCTGCCGCC 1534

Qy 25 AspaSnAlaAla 28
   ||| |||||
Db 1535 GATGCTGCGGCG 1546

RESULT 19
US-60-145-134-143
; Sequence 143, Application US/60145134
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; FILE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES THERE
; FILE REFERENCE: CL000058
; CURRENT APPLICATION NUMBER: US/60/145,134
; CURRENT FILING DATE: 1999-07-22
; NUMBER OF SEQ ID NOS: 342
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 143
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Drosophila
US-60-145-134-143

Alignment Scores:
Pred. No.: 79.7 Length: 1785
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 53 Caps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-145-134-143 (1-1785)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1477 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGACAGGATGCCGCGCTGCCGCC 1536

Qy 25 AspaSnAlaAla 28
   ||| |||||
Db 1537 GATGCTGCGGCG 1546

RESULT 20
US-60-173-464-14401/c
; Sequence 14401, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; FILE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14401
```

```
; LENGTH: 3046
; TYPE: DNA
; ORGANISM: Drosophila
US-60-173-464-14401

Alignment Scores:
Pred. No.: 148 Length: 3046
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 56 Caps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-173-464-14401 (1-3046)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1014 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGACAGGATGCCGCGCTGCCGCC 955

Qy 25 AspaSnAlaAla 28
   ||| |||||
Db 954 GATGCTGCGGCG 943

RESULT 21
US-60-167-217-17590/c
; Sequence 17590, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; FILE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17590
; LENGTH: 3049
; TYPE: DNA
; ORGANISM: Drosophila
US-60-167-217-17590

Alignment Scores:
Pred. No.: 148 Length: 3049
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 55 Caps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-167-217-17590 (1-3049)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1017 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGACAGGATGCCGCGCTGCCGCC 958

Qy 25 AspaSnAlaAla 28
   ||| |||||
Db 957 GATGCTGCGGCG 946

RESULT 22
US-09-614-150-17531/c
; Sequence 17531, Application US/09614150
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; FILE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
```

## \* TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/993.002A  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 4235:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1239 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...1239

US-08-993-002A-4235

Alignment Scores:		
Pred. No.:	15.6	Length: 1239
Score:	57.00	Matches: 12
Percent Similarity:	69.57%	Conservative: 4
Best Local Similarity:	52.17%	Mismatches: 7
Query Match:	44.19%	Indels: 0
DB:	13	Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-993-002A-4235 (1-1239)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
Db 145 ATACAGGCCCTACAGGACCAATTGACGCTTAGATTCTCAAGMAAAGTCGTTAGCAAA 204

Qy 24 AlaAspAsn 26

Db 205 TGGGATAAC 213

RESULT 17

US-08-621-425-49  
Sequence 49, Application US/08621425  
GENERAL INFORMATION:

APPLICANT: Douglas Smith  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 495  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, Suite 510

CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 8mm cartridge tape  
COMPUTER: SPARC station LX  
OPERATING SYSTEM: sunOS4  
SOFTWARE: tar  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/621.425  
FILING DATE: 25-MARCH-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/561.469  
FILING DATE: 17-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487.032  
FILING DATE: 07-JUNE-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-001CP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3497 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...3497

US-08-621-425-49

Alignment Scores:		
Pred. No.:	51.9	Length: 3497
Score:	57.00	Matches: 12
Percent Similarity:	69.57%	Conservative: 4
Best Local Similarity:	52.17%	Mismatches: 7
Query Match:	44.19%	Indels: 0
DB:	10	Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-621-425-49 (1-3497)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
Db 1528 ATACAGGCCCTACAGGACCAATTGACGCTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 1587

Qy 24 AlaAspAsn 26

Db 1588 TGGGATAAC 1596

RESULT 18

US-60-142-845-289  
Sequence 289, Application US/60142845  
GENERAL INFORMATION:

APPLICANT: Kerlavage, Anthony  
TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: CLO00048  
CURRENT APPLICATION NUMBER: US/60/142.845  
CURRENT FILING DATE: 1999-07-08  
NUMBER OF SEQ ID NOS: 704  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 289

```
; MEDIUM TYPE:
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA: PCT/US97/19575
; FILING DATE:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/759,739
; FILING DATE: 06-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/891,928
; FILING DATE: 14-JULY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP10PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...1239
PCT-US97-19575-51
```

```
Alignment Scores:
Pred. No.: 15.6 Length: 1239
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 1 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x PCT-US97-19575-51 (1-1239)

```
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 145 ATACAAGCCCTACAGGAGCAAAATGACGCTTTAGATTCTCAAGAAAAGTCGTTAGCAAA 204
Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213
```

```
RESULT 15
US-08-759-739-250
; Sequence 250, Application US/08759739
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI VACCINE COMPOSITIONS THERE
; NUMBER OF SEQUENCES: 608
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA: US/08/759,739
; FILING DATE:
; APPLICATION NUMBER: US 08/487,032
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APRIL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 250:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...1239
US-08-759-739-250
```

```
Alignment Scores:
Pred. No.: 15.6 Length: 1239
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 11 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-759-739-250 (1-1239)

```
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 145 ATACAAGCCCTACAGGAGCAAAATGACGCTTTAGATTCTCAAGAAAAGTCGTTAGCAAA 204
Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213
```

```
RESULT 16
US-08-993-002A-4235
; Sequence 4235, Application US/08993002A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
```

APPLICATION NUMBER: US/08/759,739  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,032  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/561,469  
FILING DATE: 17-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/630,405  
FILING DATE: 01-APRIL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/660,742  
FILING DATE: 06-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/736,791  
FILING DATE: 25-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/739,150  
FILING DATE: 28-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-001CP8  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 92:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 666 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1...666  
US-08-759-739-92

Alignment Scores:  
Pred. No.: 7.59 Length: 666  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 11 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-759-739-92 (1-666)  
QY 4 LeuGlulAlaLeuAlaaspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
Db 229 ATCAAGCCCTACAGGAGCAATTGACGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 288

QY 24 AlaAspAsn 26  
Db 289 TGGGATAAC 297

RESULT 13  
US-08-993-002A-4234  
Sequence 4234, Application US/08993002A  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street

CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/993,002A  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 4234:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 666 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1...666  
US-08-993-002A-4234

Alignment Scores:  
Pred. No.: 7.59 Length: 666  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 13 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-993-002A-4234 (1-666)  
QY 4 LeuGlulAlaLeuAlaaspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
Db 229 ATCAAGCCCTACAGGAGCAATTGACGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 288

QY 24 AlaAspAsn 26  
Db 289 TGGGATAAC 297

RESULT 14  
PCT-US97-19575-51  
Sequence 51, Application PC/TUS9719575  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI AND  
TITLE OF INVENTION: VACCINE COMPOSITIONS THEREOF  
NUMBER OF SEQUENCES: 208  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:

```
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APRIL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...576
US-08-759-739-43

Alignment Scores:
Pred. No.: 6.41 Length: 576
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 11 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-759-739-43 (1-576)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 145 ATACAGCCCTACAGGCAAAATTGACGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAA 204

Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

RESULT 11
US-08-993-002A-4232
; Sequence 4232, Application US/08993002A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993.002A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4232:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...576
US-08-993-002A-4232

Alignment Scores:
Pred. No.: 6.41 Length: 576
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 13 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-993-002A-4232 (1-576)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 145 ATACAGCCCTACAGGCAAAATTGACGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAA 204

Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

RESULT 12
US-08-759-739-92
; Sequence 92, Application US/08759739
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI VACCINE COMPOSITIONS THERE
; NUMBER OF SEQUENCES: 608
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
```

;; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR  
;; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

;; NUMBER OF SEQUENCES: 880

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: LAHIVE & COCKFIELD

;; STREET: 60 State Street

;; CITY: Boston

;; STATE: Massachusetts

;; COUNTRY: USA

;; ZIP: 02109-1875

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/487,032A

;; FILING DATE: 07-JUNE-1995

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Mandragouras, Amy E.

;; REGISTRATION NUMBER: 36,207

;; REFERENCE/DOCKET NUMBER: GTN-001

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (617)227-7400

;; TELEFAX: (617)227-5941

;; INFORMATION FOR SEQ ID NO: 282:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 576 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: double

;; TOPOLOGY: circular

;; MOLECULE TYPE: DNA (genomic)

;; HYPOTHETICAL: NO

;; ANTI-SENSE: NO

;; ORIGINAL SOURCE:

;; ORGANISM: Helicobacter pylori

;; US-08-487-032A-282

Alignment Scores:

Pred. No.: 6.41 Length: 576  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 8 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-487-032A-282 (1-576)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23

Db 145 ATACAAGCCCTACAGGAGCAAAATTGACGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 204

Qy 24 AlaAspAsn 26

Db 205 TGGGATAAC 213

RESULT 9

US-08-561-469A-282

;; Sequence 282, Application US/08561469A

;; GENERAL INFORMATION:

;; APPLICANT: DOUGLAS SMITH

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

;; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR

;; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

;; NUMBER OF SEQUENCES: 994

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: LAHIVE & COCKFIELD

;; STREET: 60 State Street, Suite 510

;; CITY: Boston

;; STATE: Massachusetts

;; COUNTRY: USA

;; ZIP: 02109-1875

;; COMPUTER READABLE FORM:

;; APPLICATION NUMBER: US 08/487,032

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/561,469A

;; FILING DATE: 17-NOV-1995

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/487,032

;; FILING DATE: 07-JUNE-1995

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Mandragouras, Amy E.

;; REGISTRATION NUMBER: 36,207

;; REFERENCE/DOCKET NUMBER: GTN-001CP

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (617)227-7400

;; TELEFAX: (617)227-5941

;; INFORMATION FOR SEQ ID NO: 282:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 576 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: double

;; TOPOLOGY: circular

;; MOLECULE TYPE: DNA (genomic)

;; HYPOTHETICAL: NO

;; ANTI-SENSE: NO

;; ORIGINAL SOURCE:

;; ORGANISM: Helicobacter pylori

;; US-08-561-469A-282

Alignment Scores:

Pred. No.: 6.41 Length: 576  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 9 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-561-469A-282 (1-576)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23

Db 145 ATACAAGCCCTACAGGAGCAAAATTGACGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 204

Qy 24 AlaAspAsn 26

Db 205 TGGGATAAC 213

RESULT 10

US-08-759-739-43

;; Sequence 43, Application US/08759739

;; GENERAL INFORMATION:

;; APPLICANT: DOUGLAS SMITH ET AL

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

;; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI VACCINE COMPOSITIONS THERE

;; NUMBER OF SEQUENCES: 608

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: LAHIVE & COCKFIELD

;; STREET: 60 State Street, Suite 510

;; CITY: Boston

;; STATE: Massachusetts

;; COUNTRY: USA

;; ZIP: 02109-1875

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: CD-ROM ISO9660

;; COMPUTER:

;; OPERATING SYSTEM:

;; SOFTWARE:

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/759,739

;; FILING DATE:

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/487,032

; APPLICANT: Rasmussen, Magnus  
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA 2M BINDING PROTEIN  
; FILE REFERENCE: 100084.415US / N.75312B  
; CURRENT APPLICATION NUMBER: US/09/847,539A  
; CURRENT FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 764  
; TYPE: DNA  
; ORGANISM: Streptococcus pyogenes  
US-09-847-539A-27

Alignment Scores:  
Pred. No.: 2,25e-12 Length: 764  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 32 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-27 (1-764)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
|||||  
Db 333 TCAGATGCCCTTAGAGCATTTAGCGGATCAACAGACGCTTTACAAATCAGAAGAAGCTGCG 392  
QY 21 ValValLysAlaAspAsnAlaAla 28  
|||||  
Db 393 GTTGTAAAGCGGATAACGCTGCT 416

## RESULT 5

US-09-847-539A-13  
; Sequence 13, Application US/09847539A  
; GENERAL INFORMATION:  
; APPLICANT: Bjorck, Lars H  
; APPLICANT: Rasmussen, Magnus  
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA 2M BINDING PROTEIN  
; FILE REFERENCE: 100084.415US / N.75312B  
; CURRENT APPLICATION NUMBER: US/09/847,539A  
; CURRENT FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 777  
; TYPE: DNA  
; ORGANISM: Streptococcus pyogenes  
US-09-847-539A-13

Alignment Scores:  
Pred. No.: 2,29e-12 Length: 777  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 32 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-13 (1-777)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
|||||  
Db 319 ACTGACGCCTTAGAGCATTTGCGGATCAACAGACGCTTTACAAATCAGAAGAAGCTGCG 378  
QY 21 ValValLysAlaAspAsnAlaAla 28  
|||||  
Db 379 GTTGTAAAGCGGATAACGCTGCT 402

## RESULT 6

US-09-847-539A-15  
; Sequence 15, Application US/09847539A  
; GENERAL INFORMATION:  
; APPLICANT: Bjorck, Lars H  
; APPLICANT: Rasmussen, Magnus

; TITLE OF INVENTION: STREPTOCOCCAL ALPHA 2M BINDING PROTEIN  
; FILE REFERENCE: 100084.415US / N.75312B  
; CURRENT APPLICATION NUMBER: US/09/847,539A  
; CURRENT FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 853  
; TYPE: DNA  
; ORGANISM: Streptococcus pyogenes  
US-09-847-539A-15

Alignment Scores:  
Pred. No.: 2,56e-12 Length: 853  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 32 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-15 (1-853)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
|||||  
Db 556 ACTGACGCCTTAGAGCATTTGCGGATCAACAGACGCTTTACAAATCAGAAGAAGCTGCG 615  
QY 21 ValValLysAlaAspAsnAlaAla 28  
|||||  
Db 616 GTTGTAAAGCGGATAACGCTGCT 639

## RESULT 7

US-09-489-039A-318  
; Sequence 318, Application US/09489039A  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 318  
; LENGTH: 828  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-318

Alignment Scores:  
Pred. No.: 6.53 Length: 828  
Score: 50.00 Matches: 12  
Percent Similarity: 75.00% Conservative: 6  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 44.96% Indels: 0  
DB: 18 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-489-039A-318 (1-828)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 21  
|||||  
Db 157 GACCGTCTCAGCGGCTCGCGACAGCTGCAGGCGCGCTACGCCATCCAGCAAGCATC 216  
QY 22 ValLysAlaAsp 25  
|||||  
Db 217 CTCAGAGCGGAT 228

## RESULT 8

US-08-487-032A-282  
; Sequence 282, Application US/08487032A  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:20:12 : Search time 553.112 Seconds:  
(without alignments)  
1095.215 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALALADOTDALOSEAAVVKADNA 28

Scoring table: BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 21979536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -  
-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_092549\_3486/app\_query.fasta\_1.526  
-DB=Pending\_Patents\_NA\_Main -QFMT=fastap -SUFFIX=trmp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptlo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539\_@CGN\_1\_1\_1201\_@runat\_10102002\_092549\_3486 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Pending\_Patents\_NA\_Main:\*

1:	/cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2:	/cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3:	/cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4:	/cgn2_6/ptodata/2/pna/US08_COMB.seq:*
5:	/cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6:	/cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7:	/cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8:	/cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9:	/cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10:	/cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11:	/cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12:	/cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13:	/cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14:	/cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15:	/cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16:	/cgn2_6/ptodata/2/pna/US092_COMB.seq:*
17:	/cgn2_6/ptodata/2/pna/US093_COMB.seq:*
18:	/cgn2_6/ptodata/2/pna/US094_COMB.seq:*
19:	/cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
20:	/cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
21:	/cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
22:	/cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
23:	/cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
24:	/cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
25:	/cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
26:	/cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
27:	/cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
28:	/cgn2_6/ptodata/2/pna/US097A_COMB.seq:*

29:	/cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
30:	/cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
31:	/cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
32:	/cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
33:	/cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
34:	/cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
35:	/cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
36:	/cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
37:	/cgn2_6/ptodata/2/pna/US100_COMB.seq:*
38:	/cgn2_6/ptodata/2/pna/US101_COMB.seq:*
39:	/cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
40:	/cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
41:	/cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
42:	/cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
43:	/cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
44:	/cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
45:	/cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
46:	/cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
47:	/cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
48:	/cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
49:	/cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
50:	/cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
51:	/cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
52:	/cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
53:	/cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
54:	/cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
55:	/cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
56:	/cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
57:	/cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
58:	/cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
59:	/cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
60:	/cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
61:	/cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
62:	/cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
63:	/cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
64:	/cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
65:	/cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
66:	/cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
67:	/cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
68:	/cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
69:	/cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
70:	/cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
71:	/cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
72:	/cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
73:	/cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
74:	/cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
75:	/cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	129	100.0	469	32	US-09-847-539A-14
2	129	100.0	504	32	US-09-847-539A-16
3	129	100.0	654	32	US-09-847-539A-12
4	129	100.0	764	32	US-09-847-539A-27
5	129	100.0	777	32	US-09-847-539A-13
6	129	100.0	853	32	US-09-847-539A-15
7	58	45.0	828	18	US-09-489-039A-318
8	57	44.2	576	8	US-08-487-032A-282
9	57	44.2	576	9	US-08-561-469A-282
10	57	44.2	576	11	US-08-759-739-43
11	57	44.2	576	13	US-08-993-002A-4232
12	57	44.2	666	11	US-08-759-739-92
13	57	44.2	666	13	US-08-993-002A-4234
14	57	44.2	1239	1	PCT-US97-19575-51
15	57	44.2	1239	11	US-08-759-739-250
16	57	44.2	1239	13	US-08-993-002A-4235

```

REFERENCE
AUTHORS      2 (bases 1 to 848)
              Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
              Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
              de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
              Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
              Saurin,W., Tekalia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
              Wincker,P. and Weissenbach,J.
TITLE        Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
              yeast species for molecular evolution studies(1)
JOURNAL      FEBS Lett. 487 (1), 3-12 (2000)
PUBMED       11152876
AUTHORS      3 (bases 1 to 848)
              Genoscope.
JOURNAL      Direct Submission
              Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
              2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
              seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT      This STS is part of a random genomic sequencing program of thirteen
              yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
              exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
              Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
              lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
              angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
              Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
              5 kb were prepared and both extremities were sequenced. See
              keywords for description of this sequence and for the sequence of
              the other extremity of this insert.
FEATURES
SOURCE       1..848
              Location/Qualifiers
              /organism="Saccharomyces bayanus"
              /strain="CUIB 533"
              /variety="uvarum"
              /db_xref="taxon:4931"
              /clone="AS0AA018G01"
              /clone_lib="AS0AA"
              /note="end : 17"
              <20..>180
misc_feature  /note="similar to Saccharomyces cerevisiae ORF YOL110w [
              SHR5 ; RAS suppressor ]
              1 putative frameshift(s)"
              /evidence=not_experimental
              <577..>846
misc_feature  /note="similar to Saccharomyces cerevisiae ORF YOL109w [
              ZEO1 ; zeocin resistance ]"
              /evidence=not_experimental
BASE COUNT   211 a 205 c 189 g 240 t
ORIGIN
Alignment Scores:
Pred. No.:    0.0504      Length:      848
Score:         88.00      Matches:     21
Percent Similarity: 58.82%  Conservative: 9
Best Local Similarity: 41.18%  Mismatches: 21
Query Match:    11.33%      Indels:    0
DB:            11          Gaps:      0

US-09-847-539A-6 (1-159) x CNS0618W (1-848)
QY   81 LysAlaAspAsnAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
DB   595 AAAGCTGAAACCTGGCTGAAGACCTCCAAACAAAGCTGGATGAAACCAAGGAGTCCTTA 654
QY   101 GlnSerGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
DB   655 CAACACAGAGCGCCGAGTGAAGGACACACAGCCCAACTGCTGTCGACGACTTGAAGAG 714
QY   121 AlaAlaThrProIleAlaLeuAspValLysLys 131
DB   715 GAAGCTACCCAGAGCCGCAACAAAGTTAAGAAG 747

Search completed: October 13, 2002, 03:00:40
Job time : 1854.74 secs

```



LMFLCHR34\_02 200001 310000  
LMFLCHR34\_03 300001 410000  
LMFLCHR34\_04 400001 510000  
LMFLCHR34\_05 500001 610000  
LMFLCHR34\_06 600001 710000  
LMFLCHR34\_07 700001 810000  
LMFLCHR34\_08 800001 910000  
LMFLCHR34\_09 900001 1010000  
LMFLCHR34\_10 1000001 1110000  
LMFLCHR34\_11 1100001 1210000  
LMFLCHR34\_12 1200001 1310000  
LMFLCHR34\_13 1300001 1410000  
LMFLCHR34\_14 1400001 1510000  
LMFLCHR34\_15 1500001 1610000  
LMFLCHR34\_16 1600001 1710000  
LMFLCHR34\_17 1700001 1720777

Continuation (17 of 18) of LMFLCHR34 from base 1600001 (AL499623 Leishmania major chromo

#### Alignment Scores:

Pred. No.: 7.07 Length: 110000  
Score: 89.00 Matches: 25  
Percent Similarity: 54.43% Conservative: 18  
Best Local Similarity: 31.65% Mismatches: 36  
Query Match: 11.45% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539a-6 (1-159) x LMFLCHR34\_16 (1-110000)

QY 44 ALaIleGluAspLysGluAlaThrThrAlaIleGluAlaAAlaSerSerAspAlaLeuGlu 63  
Db 100106 GCGCGTGGCCAGCGGCGTCTGCTGCTGGAGGAAAGCGCGGCTGCCGAGAGC 100165  
QY 64 ALaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValysAlaAsp 83  
Db 100166 GAGCTGGCTGCTGCGCGGCGGCGATACGCGCTGCTGAGGCGGCGCTGCCGAGCGGAC 100225  
QY 84 AnAlaIaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 103  
Db 100226 GCGGAGCATGCCGCGGCTGCTGCTGCGGAGGCGGCGTGTGCTGCCGAGCGTGAG 100285  
QY 104 GluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTlpGluLysAlaAla 122  
Db 100286 GCGCGTGTGCTGGAGCGGAGGCGGCGGCTCCGAGCGGAGCTGGCTGCTGCG 100342

RESULT 39  
AP002556  
LOCUS AP002556 291136 bp DNA linear BCT 07-MAR-2001  
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 7/20.  
ACCESSION AP002556 BA000007  
VERSION AP002556.1 GI:13361156  
KEYWORDS  
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub\_strain:RIMD 0509952)  
DNA.

ORGANISM Escherichia coli O157:H7  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.

REFERENCE 1 (sites)  
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yanamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.  
TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak  
Genes Genet. Syst. 74 (5), 227-239 (1999)

JOURNAL  
MEDLINE  
AUTHORS  
2 (sites)  
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.  
TITLE Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
JOURNAL

COMMENT  
FEATURES  
source

gene  
CDS

gene  
CDS

Syst. Appl. Microbiol. 23 (3), 315-324 (2000)  
20557356

3 (sites)  
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.  
TITLE Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak  
Gene 258 (1-2), 127-139 (2000)

Gene 258 (1-2), 127-139 (2000)  
20564182  
4 (sites)  
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.  
TITLE Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12  
DNA Res. 8 (1), 11-22 (2001)

21156231  
5 (bases 1 to 291136)  
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.  
TITLE Direct Submission  
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, Tel:81-6-6879-8365, URL:http://www.gen-info.osaka-u.ac.jp/, Fax:81-6-6879-2047)  
genome project.

Location/Qualifiers  
1. 291136  
/organism="Escherichia coli O157:H7"  
/strain="O157:H7"  
/sub\_strain="RIMD 0509952"  
/db\_xref="taxon:83334"  
126..380  
/gene="ECs1692"  
126..380  
/gene="ECs1692"  
/note="similar to YMG-ECOLI gi11787445 percent identity 97 in 84 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="hypothetical protein"  
/protein\_id="BAB35115.1"  
/db\_xref="GI:13361157"  
/translation="MGIIATIFGLIAGILAKLIMPGRDGGGFELTLCILGIVGAVVG WLATMFGIGSGISGFNLHSLFVAVVGVAILVLGVFLRLRRE"  
complement(430..2400)  
/gene="ECs1693"  
complement(430..2400)  
/gene="ECs1693"  
/note="probable TonB dependent outer membrane receptor, similar to TonB dependent outer membrane receptor PrrA [Escherichia coli CFT073] gi13661477|gb|AAC61709.1| percent identity 97 in 656 aa"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="putative TonB dependent outer membrane receptor"  
/protein\_id="BAB35116.1"  
/db\_xref="GI:13361158"

/translation="MRUKKHYLCTALSIAFTQQAAVAQESLSLTVMSSPVSSTTTTVD OPTMKALDKONVAALSWFVYLOKSGSRNEQVKRGDSRVFYFDGVTIYVP WSRDNLARILTNLNGAVEVSGYSSLIQGPNGMGAINITQPTKPLEASLAGYRQG YRSODNAYDMHSAFSSSELGYLOVSGSLQKDFLGLPHGVNNDIAGKHGKMLNSSA DDKRGIVKLGCTPRENDEYLTITYIKDGEKDNPPYSGNSGOKSRWQWPEYDKSFYV QGTQLNDRFTLKRLYRDNFTENTIMYNSLADLNKRGKSYSHYSDYSDGAGLQALAD VRENDLLTFAVNWKDDVHREKGAHAYDRYEDRTWSLASLEYQWAAADNVDDVAGISV

NLSMWAVLQOMDGRVLYIAGIGAGIENTPAGMOSOVLLAADRMIANPANGTKPM  
FYVGDQDIPNNRVFLKYLTAFTITSGGNPPAFSLPDRGLTAKNADISGNVANSGTL  
NNVTINENCRVLGKUSAMQIEGLDKVKAAPRDSRAPRPSGTITTVRVYDDQFDF  
RQIVTAVAFSGAKHEKHEHTDIYSSCLIVRKNGAEIYRNTALDNTLIYVYDDMPAG  
HGHMTLEFSVAWLNNMYPTASISDLLVVVKKATAGITIS"

gene

CDS

/gene="26030"  
/complement(5588..6943)  
/gene="26030"  
/function="putative structure; Structural component (Phage  
or Phage Related)"  
/note="Residues 61 to 451 of 451 are 96.93 pct identical  
to residues 1 to 391 of 391 from Genpept121:  
dbj|BAB19563.1| (AP000400) host specificity protein J  
[Escherichia coli O157:H7]"  
/codon\_start=1  
/transl\_table=11  
/product="putative tail component of cryptic prophage  
CP-933p"

/protein\_id="AAK16946.1"  
/db\_xref="GI:13259575"  
/translation="MPLPVLTRIIIGHERKPRYLIPGFSIAWOKVCVSCFSGEYPT  
SVNKKWRLIYSGEGCMKGKGAHTPREAKDNLSKTMSSVIDAIGEGPIEGPVKGL  
QSLVYKNTDLTDGPNVTHGVAVRAGEQOTPEGFESSGAETGLGVEVTKAPV  
TRTITSANTLRTVFGVQSLETTSKGDRNHSVRLLIQLORNGNMVTEKDVTLNCK  
TTSQFLASVILNLPENIRKRETAISTDQONKTLSSYTEIIDYKQCPNTA  
TYQLQVDFQGGQQTNYHIRIIQVPSYNDPEKRTYSGIWDGSLRPAYSNNPWW  
CLWMDLTHPRYGGKRLGAADVKNALYAIQYCDQTVDFGFGTEPRMTFNALYSQ  
PRAWVDLSFCSAMRCMPVYNGQTLTFVQDRPSDVVWPTPAYVMMWMITAWGFATASA  
P"

gene

CDS

/complement(7105..7785)  
/gene="26031"  
/complement(7105..7785)  
/gene="26031"  
/function="putative factor; DNA packaging, phage assembly  
(Phage or Phage Related)"  
/note="Residues 1 to 224 of 226 are 92.85 pct identical to  
residues 1 to 224 of 225 from Genpept121: dbj|BAB19564.1|  
(AP000400) tail assembly protein I [Escherichia coli  
O157:H7]"  
/codon\_start=1  
/transl\_table=11

/product="putative tail assembly protein of cryptic  
prophage CP-933p"  
/protein\_id="AAK16947.1"  
/db\_xref="GI:13259576"

/translation="MATTNFAFLASPLARICLHGLQRFRRRLSYLVNTAAIRAL  
SLQVQFRRQMEGHTYIRIAGDYTAPEAVYARLHEOLGEGTVIHVPLRAGAKGGL  
QIVLGAATAVGSFETAGASMLAGSALAGPSATMFLSLGASMLTGLGVAOMLAPKA  
KTPDVRATONGRQNTYFSLDNWIAOCNPMYPYGMVLSGRRISQDISTRBGGGCT  
VVVVGROG"

gene

CDS

/complement(7683..8357)  
/gene="26032"  
/complement(7683..8357)  
/gene="26032"  
/function="putative factor; DNA packaging, phage assembly  
(Phage or Phage Related)"  
/note="Residues 1 to 223 of 224 are 92.37 pct identical to  
residues 24 to 246 of 247 from Genpept121: dbj|BAB19565.1|  
(AP000400) tail assembly protein K [Escherichia coli  
O157:H7]"

Alignment Scores:

Pred. No.: 0.496 Length: 9331  
Score: 89.00 Matches: 27  
Percent Similarity: 50.00% Conservative: 7  
Best Local Similarity: 39.71% Mismatches: 34  
Query Match: 11.45% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x AE006458 (1-9331)

Qy 50 AlathrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69

Db 2031 GCTGACACTTCACAGGGGATGCATCGGAGTCAGCCGCGGCGGAGAAAGTGCAGCC 1972  
Qy 70 AspAlaLeuGlnSerGluGluAlaAlaValLysAlaAspAsnAlaAlaSerAspAla 89  
Db 1971 GCTGCAAGACAGTCACAGGAGGCGTCTCGTCGCGCCTCTCGCGCGCTCAAAAAGCC 1912  
Qy 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln 109  
Db 1911 AGTCAGTCATTACAAAGTCACACAGATGCTGAGTTGTCAANAAGAGCGGAGCAAGTGC 1852  
Qy 110 SerAspAsnAlaAlaSerAspAla 117  
Db 1851 GCCGTTAATGCAGCCAGGATGCA 1828

RESULT 37

LMFLCHR34\_17

WPCOMMENT

Sequence split into 18 fragments LOCUS LMFLCHR34 Accession AL499623

Fragment Name	Begin	End
LMFLCHR34_00	1	110000
LMFLCHR34_01	100001	210000
LMFLCHR34_02	200001	310000
LMFLCHR34_03	300001	410000
LMFLCHR34_04	400001	510000
LMFLCHR34_05	500001	610000
LMFLCHR34_06	600001	710000
LMFLCHR34_07	700001	810000
LMFLCHR34_08	800001	910000
LMFLCHR34_09	900001	1010000
LMFLCHR34_10	1000001	1110000
LMFLCHR34_11	1100001	1210000
LMFLCHR34_12	1200001	1310000
LMFLCHR34_13	1300001	1410000
LMFLCHR34_14	1400001	1510000
LMFLCHR34_15	1500001	1610000
LMFLCHR34_16	1600001	1710000
LMFLCHR34_17	1700001	1720777

Continuation (18 of 18) of LMFLCHR34 from base 1700001 (AL499623 Leishmania major chr

Alignment Scores:

Pred. No.: 1.17 Length: 20777  
Score: 89.00 Matches: 25  
Percent Similarity: 54.43% Conservative: 18  
Best Local Similarity: 31.65% Mismatches: 36  
Query Match: 11.45% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6 (1-159) x LMFLCHR34\_17 (1-20777)

Qy 44 AlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAspAlaLeuGlu 63  
Db 106 GCCGCTGCCAGCGGAGGCTGCTGCTGCGTGGAGGCAAGAGCGGCGCTCCGAGAGC 165  
Qy 64 AlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLysAlaAsp 83  
Db 166 GAGCTGGCTGCTGCGCGGAGCGATACGCGCTGCTGAGCGCGGCTGGCGGAGCGGAC 225  
Qy 84 AsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 103  
Db 226 GCCGAGCATGCCCGCTGCTGCTGCGTGGAGCGGAGGCGTGTGCTGCCGAGCGTGAG 285  
Qy 104 GluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAla 122  
Db 286 GCCGCTGCTGCTGCGTGGAGCGGAGAGCGCGCTCCGAGAGCGAGCTGCTGCTGCGC 342

RESULT 38

LMFLCHR34\_16

WPCOMMENT

Sequence split into 18 fragments LOCUS LMFLCHR34 Accession AL499623

Fragment Name	Begin	End
LMFLCHR34_00	1	110000
LMFLCHR34_01	100001	210000







```

Oy 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125
    ||| :::: |||||| ::||| ::||| |||||
Db 123840 GAGAGGCGCGCCCAAGAGCGCCGAGAGCTGCCCAAAAGGTGCGGAAGAGCC 123781
Oy 126 AlaLeu 127
    ||||||
Db 123780 GCTTTA 123775

RESULT 34
AC020883
LOCUS AC020883 269726 bp DNA linear HTG 15-JUL-2000
DEFINITION Mus musculus clone RP23-441L2, WORKING DRAFT SEQUENCE, 56 unordered
pieces.
ACCESSION AC020883
VERSION AC020883.3 GI:9211213
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 269726)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Mouse
REFERENCE 2 (bases 1 to 269726)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 15, 2000 this sequence version replaced gi:5980210.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1906482
Center clone name: RPCI-23_441L2
-----
Summary Statistics
Consensus quality: 205453 bases at least Q40
Consensus quality: 224960 bases at least Q30
Consensus quality: 223624 bases at least Q20
Estimated insert size: 240000; pulse field gel estimation
Estimated insert size: 264226; sum-of-contigs estimation
Quality coverage: 8.93 in Q20 bases; pulse field gel estimation
Quality coverage: 8.11 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
1 1018: contig of 1018 bp in length
* 1019: gap of unknown length
* 1119: contig of 1030 bp in length
* 2149: gap of unknown length
* 3293: contig of 1044 bp in length
* 3293: gap of unknown length
* 3393: contig of 1687 bp in length
* 5080: gap of unknown length
* 5180: contig of 1196 bp in length
* 6376: gap of unknown length
* 6476: contig of 1307 bp in length
* 7783: gap of unknown length
* 7883: contig of 1098 bp in length
* 8981: gap of unknown length
* 9081: contig of 1110 bp in length
* 10191: gap of unknown length
* 10291: gap of unknown length
* 11658: contig of 1368 bp in length
* 11659: gap of unknown length

12977: contig of 1219 bp in length
13077: gap of unknown length
14244: contig of 1167 bp in length
14345: gap of unknown length
14345: contig of 1526 bp in length
15870: gap of unknown length
15970: gap of unknown length
16974: contig of 1004 bp in length
17074: gap of unknown length
18099: contig of 1025 bp in length
18199: gap of unknown length
19629: contig of 1430 bp in length
19729: gap of unknown length
21574: contig of 1845 bp in length
21674: gap of unknown length
22877: contig of 1203 bp in length
22977: gap of unknown length
24150: contig of 1173 bp in length
24250: gap of unknown length
25446: contig of 1196 bp in length
25546: gap of unknown length
27824: contig of 2278 bp in length
27924: gap of unknown length
28943: contig of 1019 bp in length
29043: gap of unknown length
30190: contig of 1147 bp in length
30290: gap of unknown length
32310: contig of 2020 bp in length
32410: gap of unknown length
33950: contig of 1540 bp in length
34050: gap of unknown length
35168: contig of 1118 bp in length
35268: gap of unknown length
36594: contig of 1326 bp in length
36694: gap of unknown length
37965: contig of 1271 bp in length
38065: gap of unknown length
39209: contig of 1144 bp in length
39309: gap of unknown length
40349: contig of 1040 bp in length
40449: gap of unknown length
41614: contig of 1165 bp in length
41714: gap of unknown length
43023: contig of 1309 bp in length
43123: gap of unknown length
44754: contig of 1631 bp in length
44854: gap of unknown length
46830: contig of 1976 bp in length
46930: gap of unknown length
48196: contig of 1266 bp in length
48296: gap of unknown length
49474: contig of 1178 bp in length
49574: gap of unknown length
50644: contig of 1070 bp in length
50744: gap of unknown length
51816: contig of 1072 bp in length
51916: gap of unknown length
53238: contig of 1322 bp in length
53338: gap of unknown length
54823: contig of 1485 bp in length
54923: gap of unknown length
57057: contig of 2134 bp in length
57157: gap of unknown length
58383: contig of 1226 bp in length
58483: gap of unknown length
60241: contig of 1758 bp in length
60341: gap of unknown length
61567: contig of 1226 bp in length
61667: gap of unknown length
63287: contig of 1620 bp in length
63387: gap of unknown length
64876: contig of 1489 bp in length
64976: gap of unknown length
66343: contig of 1367 bp in length

```

```
/protein_id="AAG22139.1"
/db_xref="GI:10726439"
/translation="MATSKWESDEEIOYFREYLRIPSVHPDPVAPCVFELRRQANLM
DLPKMYYPANEQNPVVLTKGLNPELPSILLNSHDVVPFENWTHPFGADIDE
EGRIIPARKQDMKSGVGHQHLAAEALRKSGAKFRKTHISEVADEEMGGRYGMRFPVP
TEDRALNVGFMDEGLASDEQIPLFYAERAVRVYFNISGTAGHGSLLPNTAGEK
LNYTVGKMEFRQRVQLQNLNPELVIGDVTINLTKLGGVQSVNVPLLMVCFDCR
LALVDPEERFANLHKCADVGGGIEITYEOKPKVPPTADIGSNPFWLAFKKATDEM
HISIKPQIFGCTDSRYIRAVGIPALGFSPNNTPVLLHDHDEFIQADIIYLRGQIFQ
KIISNIANV"
complement(join(<32971..33155,33431..34088,34241..>34261))
/gene="CG14688"
/product="CT34474"
/db_xref="FLYBASE:FBan0014688"
/db_xref="FLYBASE:FBgn0037819"
/evidence-not_experimental
complement(<32971..>34261)
/gene="CG14688"
/map="86C2-86C2"
/db_xref="FLYBASE:FBan0014688"
/db_xref="FLYBASE:FBgn0037819"
/evidence-not_experimental
complement(join(<32971..33155,33431..34088,34241..>34261))
/gene="CG14688"
/codon_start=1
/db_xref="FLYBASE:FBan0014688"
/evidence-not_experimental
/protein_id="AAF54534.2"
/db_xref="GI:10726440"
/translation="MQSVNRSLNENGIYVIEDFLTAEVDLSYQNGRACLDAPQNNR
KIFSTIQEDAQGRLEFMESGDVURFFFKAGVGDQQLLVDMIALNKYGHALHVE
HPAFNATFNRVREIQWLKNYPKAVQCSMYIYKNGVGVGVPHODSWFLHDTNPS
AVGFWLALEDCTIQGLQFTKSHKSGVHRRYLRNPKDKASELMYVDRAAPIYPOSS
FTPMQVSKGTCILIHGNVHKSEPNRSQKSHAYTFHVIETENNKYSEDNWLQAPKD
KPFVLPFERA"
join(42080..42386,44345..44848,45327..>45631)
/gene="CG14681"
/product="CT34467"
/db_xref="FLYBASE:FBan0014681"
/db_xref="FLYBASE:FBgn0037820"
<42080..>45631
/gene="CG14681"
/map="86C2-86C2"
/db_xref="FLYBASE:FBan0014681"
/db_xref="FLYBASE:FBgn0037820"
join(42326..42386,44345..44848,45327..>45631)
/gene="CG14681"
/note="CG14681 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014681"
/db_xref="FLYBASE:FBgn0037820"
/protein_id="AAF54535.1"
/db_xref="GI:7299343"
/translation="MLAMKDKPMLLLFCLLALSLASFQDAAPYYGTYKIGALTRLH
HGVSQDYYVDSRTIFIKFNKYNDGEAPAFYVGNRTARPSNEGAARLDERGGTASLT
RRYKNDVTLSVPGKTLRDIKFSVWCDEFAVNFQDVSTIPNLDFPRPKISALRGV
HGVSNDIVLADQATLVLVLPNPSYDGEAPDAKFYGRGQRPTSDGLRIPDENGKPNLR
RYERTIVLTPLEDTIFDIGHFGWCEAFVDFGHVRLPEGLNVPVPSLKMGLISPOV
KKNASQILSELQR"
join(<46218..46232,46434..46533,47430..48206,48265..>48458,
48515..>51130)
/gene="CG14682"
/product="CT34468"
/db_xref="FLYBASE:FBan0014682"
/db_xref="FLYBASE:FBgn0037821"
/evidence-not_experimental
<46218..>51130
/gene="CG14682"
/map="86C2-86C2"
/db_xref="FLYBASE:FBan0014682"
/db_xref="FLYBASE:FBgn0037821"
```

```
/evidence-not_experimental
join(46218..46232,46434..46533,47430..48206,48265..>48458,
48515..>51130)
/gene="CG14682"
/note="CG14682 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014682"
/db_xref="FLYBASE:FBgn0037821"
/evidence-not_experimental
/protein_id="AAF54536.1"
/db_xref="GI:7299344"
/translation="MOTFRSLNCEVLYDDLAFAFVMAVAGESIVVQLVAKLEPNHYM
SEGISPNKNIISOMIGADAVAAVNDPOTNGPATDVELEGKQCSGGRGACPDTKYSEK
TNSIELLNAAWNGYSIVYORSLAATDRDLDPISITGAESVVAWIGLPLNDQVEVSFII
TFYNNKHQIEFGROPKWNCPLEPGARNSSEDESDAPAAQSSTGGAGYPPAGRPN
VEPDEEYENRAEALHRQPPORQETAITQRRPVPTPKPVNSGMDIPATQCHHEPE
DCVFYAQMPTCGKHGYPATICHGVGYSWYIGLLIPELHVVRGKTYTFVVEGNNNP
DIPAKYHPFYISDDPVGYEHKREKAVIYAGLHRSRSGQVPTCGVGRJCNWTKPD
VEGPADDYQSEFAYQRTLTJLKCDAGEPCVITWKPDRTPTVYVICTHRYLGNKIIH
VHDCDSRAGLKGAAERHEIRLPAKATVAEPVPHEDYAGEASVRHETKVSANDNF
LJKHOTDLIKNNHMMTPPKUSFEITKSSEITKLISDGIIRAAEALEESLLRNPNLNP
HPNQIPNPHQKNVTPTTEISSRPEILLGTHAHTLNASPSAYSATLPSANL
KLPTILAGPHLTHIHPHLRHQHPHAPHPHVLHNLNLTANLPALAKTKTIGLSEFL
RPNAPLFFHPVKLPGRPFAPKIKVPASRPILPOQHPHLHPHPOOHVPLVLOQPSL
IVSHYKPIPGLLKPFVKEKPPLOPLAASVLLGQTELGGLNKGKRELTKGPKI
PVPYVLEPQSLONTAFNPGGKGGKQDKPKASSVSISITPTFLYKRPVTKESQE
ETASMPAVNQGFQPDVIVESGFAPVIVRTDGTGVQLPKETIIDQVAHREDPDTEIDE
VMTDTLFLAAQGGSESPFEPFIPSLDSTNATKVLNVNKEVSPASALRPSA
ALEHALPSASELIKPTLDELFAEELNELEMEPMVADVDESLETYKKDAWTTIP
IPRNTTKKPDOLLDELFGDPEEELYADELEMDRDRDAAAADITVYLPDNRKIP
DTRVPSGALYTFDCKSVSDSILVPPKLDAPDNVANHORIAQYGLTPLEQLVTRTPOF
GYVGLPQERFRTGEPQVSEYSHPPAPSRTPVPFSSSGSTIYPYSSSTGASTSVS
SSASPLSSSLSPISLTKLLKPEGRRA"
join(51689..51886,54866..54906,54969..55179,55303..55774,
56086..57923,57990..>58232)
/gene="pug"
/product="CT13530"
/db_xref="FLYBASE:FBan0004067"
<51689..>58232
/gene="pug"
/note="CG4067"
/map="86C2-86C2"
/db_xref="FLYBASE:FBan0004067"
/db_xref="FLYBASE:FBgn0020385"
join(51782..51886,52710..>54028)
```

Alignment Scores:

Pred. No.:	11.4	Length:	224795
Score:	90.00	Matches:	30
Percent Similarity:	45.10%	Conservative:	16
Best Local Similarity:	29.41%	Mismatches:	56
Query Match:	11.58%	Indels:	0
DB:	3	Gaps:	0

US-09-847-539a-6 (1-159) x AE003688 (1-224795)

QY	26	GluLysLeuAlaLeuArgAsnGluGluArgAlaIleAaspGluLeuLysGlnAlaIle	45
Db	124080	GAAGAAGCAGGCTGCTAAGGCCAAAGAGGCTGCGAGGAGGCCGCCAAATCGCTGCC	124021
QY	46	GluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu	65
Db	124020	GAAGAGGCTCTCTCTGCTGAAGCTGCCGCCCAAGAGCGAGCGCTAAGGCCCTA	123961
QY	66	AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLysAlaAspAsnAla	85
Db	123960	AAAGCGCTGAAGATGCTGCCAGAAAGCTGCCGAGAAGCGCGTCTAGCGGAGGAGGCT	123901
QY	86	AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla	105
Db	123900	GCCGCCAAAGGTCGCCGAGGAAGCTGCCCAAAAGCTGCTGAGGAAGCTCGCTAGCA	123841



Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.  
Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic *Escherichia coli* O157:H7 Sakai strain and an *Escherichia coli* K-12 strain M61655  
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)  
20557336

3 (sites)  
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.  
Complete nucleotide sequence of the prophage  $\nu$ 11-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak  
Gene 258 (1-2), 127-139 (2000)  
20564182

4 (sites)  
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.  
Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12  
DNA Res. 8 (1), 11-22 (2001)  
21156231

5 (bases 1 to 222605)  
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.  
Direct Submission  
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken-gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)  
genome project.

Location/Qualifiers  
1..222605  
/organism="Escherichia coli O157:H7"  
/strain="O157:H7"  
/sub\_strain="RIMD 0509952"  
/db\_xref="taxon:83334"  
74..607  
/gene="ECs1423"  
74..607  
/gene="ECs1423"  
/note="similar to B1045\_ECOLI g11787283 percent identity 100 in 177 aa (Conserved in E.coli K-12)."  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="putative polyprotein"  
/protein\_id="BAB34846.1"  
/db\_xref="GI:13360887"  
/translation="MKTRHVVQGDITKLVADIVYNAANPSLMGGGVGDGATHRAAGP ALLDACLVRQGGDCPTGHAVITLAGDLPAKAVVHTVGPVVRGGEQNEQLQDAYL ENSLRYANSTYSVAFPALSTCVGTGPRAAAAEIAVKTVSEFTRHALPEQVYFYCYD ENAHLYERLLTQQGDE"  
549..2030  
/gene="ECs1424"  
549..2030  
/gene="ECs1424"  
/note="similar to YMDC\_ECOLI g11787284 percent identity 99 in 493 aa (Conserved in E.coli K-12)."  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="putative synthase"  
/protein\_id="BAB34847.1"  
/db\_xref="GI:13360888"  
/translation="NMKKTPTSTKDSLPMKEMNDLPRLASVLPICSQHPGCGLFPPL EKSLDAFAARYLAEMSEHTLDVQYYIWQDDNSGRLLFSALLAAAKRGVVRVRLLLDDN

NTPLGLDILRLLDLSDHPRIEVRLENPFSPRLRLPLGYITDFSRLENRMHNSKFTVDGVV  
TIVGGRNIGDAYEGAGEEPLFSDLDVMAEPPVEDVADDFARYWYCKSVPLSQDLDV  
PVEGMADRIELPASNDAMTHYLRKMESSPFINULVDGTLPLTWAKVPLSQDLPK  
GEGKAKRHSLLQRLFDINGSPSERIDIISYFVPTQACVQALLBMRVKGVAIILTN  
SLAANDVAVHAGYARWRKKLURYVELYELAPTRQOSTLHROGITGNSGASLIUKT  
FSIDKTVFISGFNFDPSTLLNTEMGFVIESETLAQLDKRFIOSQYDAAWOLRIJDR  
WGRINWDRHAKKEIVLKEPATSFMKRVNVRASILPVEMLL"  
complement(2038..3195)  
/gene="ECs1425"  
complement(2038..3195)  
/gene="ECs1425"  
/note="similar to MDOC\_ECOLI g113025226sp1p75920 percent  
identity 99 in 385 aa (Conserved in E.coli K-12)."  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="glucans biosynthesis protein"  
/protein\_id="BAB34848.1"  
/db\_xref="GI:13360889"  
/translation="MNPVPAQREYFLDSTRAMLLGIPFHSLSYSSHTWHVNSAEP  
SLMLTLFNDFIHSFQNFVFEVLSGVESYMLFLRYPLKMKWYRVERVGIPLMTAIPLL  
TUPQIMLOVYVKGKAEWPGSLIDKYNTLAWELSHLWFLVLVVMVTTLCVWIFKRI  
RNNLESDQNTKKFSWKLVSIFLCLGIGYAVIRRTIFIVYPIILNSGTFNFIVMQTL  
FYLPFFILGALAFIFPHKALFTTPSRGCTLAALAFAYLLNORYGSGDAMWYTES  
VITMVLGLMWNVWVFSFGRHLENFQSARVTVFNASLFYLVVHHPLTLFPGAYITPHI  
TSNMLGFLCGLIFVVGCIATILYELHILRIPLLAFLEFGKPKPVKRENDKAPAR"  
3588..5123  
/gene="ECs1426"  
3588..5123  
/gene="ECs1426"  
/note="similar to MDOG\_ECOLI g11787286 percent identity  
99 in 511 aa (Conserved in E.coli K-12)."  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="periplasmic glucans biosynthesis protein"  
/protein\_id="BAB34849.1"  
/db\_xref="GI:13360890"  
/translation="MKMKRWLSAAVMLTLYTSSWAFSIDDVAKAQAQSLAGKGYEAPK  
SNLPSVFRDMKYADYQIOFNHDKRAYNNLTKFPKLEFYHOGMYEDTPVKINEVATA  
VKRIKYSPTYFTFGVDQHDKDTVDLGFAGFKVLYPINSKDNKDIIVMLGASYFVFI  
GAGVYGLSARGLAIDTALPSGEEFPRKPEWIERPKPKDKRLTIVALLDSPATGAY  
KEVWAGPGRDVTVDVQSVLYLRDKVGKLGVAPLTSMFLGPQNPQSPANNYRPELHDSNG  
LSIHAGNGEMIMRNPKHLAVSSFMENPOGFLJORGRDPSFPELDLDDYDLRPS  
AWYTPKGEWCKGSVELVEIPNDENINIVATWTDLPQPEKEMNPKYITIFSRDED  
KLHAPDNVAVQQRTRSDGVKOSNLIRQPDGTAIVVDFTGAEMKRLPDDTPVTAQTS  
ICDNGEIVESTVRYNPVTKGRLVMRVKVDKAKTKTEMRAALVYNADQTLSETWSYQLP  
ANE"  
5086..7659  
/gene="ECs1427"  
5086..7659  
/gene="ECs1427"  
/note="similar to membrane glycosyltransferase MDOH\_ECOLI  
g11787287 percent identity 100 in 847 aa but differs at  
N-ter (Conserved in E.coli K-12)."  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="membrane glycosyltransferase"  
/protein\_id="BAB34850.1"  
/db\_xref="GI:13360891"  
/translation="MKPGATSYLPNMTTEYIDAMPAAASEKAALPKTDIRAVHONLD  
ADHTRWAREDDSPQGSVKARLQWAPDLSLADQGLIKDDGROQLKAMPKAKSSMPDP  
PHRTNPVGRFNDRLGRDVTPLYRLARLTKEQESQKQRTVGTIRYITLILTLAQTV  
VATWYMKTLIPYQGWALINPMDVGDGLWVSPMQLPYMLQGIILFAVLFCWVSAG  
PMTALMGFLQLLIGRDKYSISASTVGEPLNPHRTALIMPICNDENVNRFVAGLRAW  
ESVKATNAKHEDVYLLSDSYNPDICVAEQKAMMELAEVGGEGOIYFRRRRVRKRK  
SGNIDDFCRRCWSQSYWVYVLDADSVYMGDCGLGLVRLMEANPNAGIIOSSPKASGMD  
TLYARCCQFATRYGPTFAGHVGWOLGESHYGNHAIIRVAPFTEHCALAPLPGEGS  
FAGSLSHDFVEAALMRAGWGWIAAYDLPGSYBELPNLDELDRKRWCHGNLMNF  
RFLVKGHPHVEARVFLTGVMVSYLAPLMFMFLALSTALQVVALETPOYFQPLQFL  
PWPQWRLEALALPASTVMVLLFLPKLLSILLIWCKGTKEYGFGVHTVTLLEVLFS  
VLLAPVRLFHTVVFVSAFLGWVYVNSPORDDSTSMGEAFKRGCSQLLIGLVYAVG

Qy 46 GluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu 65  
 Db 24902 GAAGAGGCTCTCCGGCTGAAGCTGCCGCCCAAAAGCAGCGGAGAGCTTA 24843  
 Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAspAsnAla 85  
 Db 24842 AAAGCCGCTGAAGATGCTGCCAGAAAGCTGCCGAAGAAGCGCTCTAGCGGAGAGCT 24783  
 Qy 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105  
 Db 24782 GCCGCCCAAAAGGTCGGGAGGAGCTGCCCAAAAGCTCTGAGGAGCTCGCTAGCA 24723  
 Qy 106 GluValValGlnSerAspAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125  
 Db 24722 GAGGAGGAGCGGCCCAAAAGCGCCGAGGAGAGCTGCCCAAAAGGCTGCCGAGAGAGCC 24663  
 Qy 126 AlaLeu 127  
 Db 24662 GCTTTA 24657  
 RESULT 31  
 AC008089/c  
 LOCUS AC008089 179669 bp DNA linear INV 24-FEB-2001  
 DEFINITION Drosophila melanogaster, chromosome 3R, region 86C-86C, BAC clone  
 BACR29B06, complete sequence.  
 ACCESSION AC008089 AC007806  
 VERSION AC008089.3 GI:13027515  
 KEYWORDS HTG  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 179669)  
 AUTHORS Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,  
 Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,  
 Rogers,Y., An,H., Balgwin,D., Bonzon,J., Beeson,K.Y., Busan,D.A.,  
 Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,  
 Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,  
 Ferrera,S., Frise,E., Galle,R.F., Garq,N.S., George,R.A.,  
 Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,  
 Ibegam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,  
 McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,  
 Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,  
 Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,  
 Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,  
 Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.  
 Sequencing of Drosophila chromosome 3R, region 86C-86C  
 Unpublished  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 2 (bases 1 to 179669)  
 Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
 Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
 Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
 Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
 Rubin,G.M.  
 Direct Submission  
 Submitted (22-JUL-1999) Drosophila Genome Center, Lawrence Berkeley  
 Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 On Feb 24, 2001 this sequence version replaced gi:5670525  
 gi:5670526.  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720  
 This sequence was assembled using end sequences from a whole genome  
 shotgun and from subclones of this BAC and its neighboring clones.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive web site (<http://www.fruitfly.org/sequence/>) or send email

to bdgpf@fruitfly.berkeley.edu.  
 FEATURES  
 source  
 1..179669  
 Location/Qualifiers  
 /organism="Drosophila melanogaster"  
 /strain="y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="3R"  
 /map="86C-86C"  
 /clone="BACR29B06 (D712)"  
 /clone\_lib="RPC1-98 (Roswell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial EcoRI in  
 pBACe3.6)"  
 BASE COUNT 52043 a 38195 c 37760 g 51671 t  
 ORIGIN  
 Alignment Scores: Length: 179669  
 Pred. No.: 8.92 Matches: 30  
 Score: 90.00  
 Percent Similarity: 45.10% Conservative: 16  
 Best Local Similarity: 29.41% Mismatches: 56  
 Query Match: 11.58% Indels: 0  
 DB: Gaps: 0  
 US-09-847-539A-6 (1-159) x AC008089 (1-179669)  
 Qy 26 GluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLysLysGlnAlaIle 45  
 Db 151758 GAAAGCAGGCTGCTGCTAAGGCCAAAGAGCTGCCGAGGAGCGCCCAAAATCGCTGCC 151699  
 Qy 46 GluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu 65  
 Db 151698 GAAGAGGCTCTCGCTGGCTGAAGCTGCCGCCCAAAAGCAGCGGAGAGCTTAAGGCCCTA 151639  
 Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 85  
 Db 151638 AAAGCCCTGAAGATGCTGCCGAGAGAGCTGCCGAGAGAGCGCTCTAGCGGAGAGGCT 151579  
 Qy 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105  
 Db 151578 GCCGCCCAAAAGGTCGGCAGGAGAGCTGCCCAAAAGCTGCTGAGGAGAGCTCGCTAGCA 151519  
 Qy 106 GluValValGlnSerAspAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125  
 Db 151518 GAGGAGCAGCGGCCCAAAAGCGCGGAGAGAGCTGCCCAAAAGGCTGCCGAGAGAGCC 151459  
 Qy 126 AlaLeu 127  
 Db 151458 GCTTTA 151453  
 RESULT 32  
 AP002555  
 LOCUS AP002555 222605 bp DNA linear BCT 07-MAR-2001  
 DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 6/20.  
 ACCESSION AP002555 BA000007  
 VERSION AP002555.1 GI:13360886  
 KEYWORDS  
 SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub\_strain:RIMD 0509952)  
 DNA.  
 ORGANISM Escherichia coli O157:H7  
 Bacteri; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.  
 REFERENCE 1 (sites)  
 AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,  
 Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,  
 Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,  
 Sasakawa,C. and Shinagawa,H.  
 Complete nucleotide sequence of the prophage VT2-Sakai carrying the  
 verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7  
 derived from the Sakai outbreak  
 Genes Genet. Syst. 74 (5), 227-239 (1999)  
 20198780  
 TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE 2 (sites)  
 AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,



```

SLRTSGKGTDAVLTTAIDTAIDTEHFRSGLPGEYTLTVRAINSYQCGGEPAITTFRI
AAPAFSRIELTCYQITAITPLHVAIVDPTQVFEFWSFKRIADIQVETAARYLGS
LWYAAASINLPGHDYFYFIRSYNTVGSFAVAGVSDADGCVDFDKTGKIGKTHL
QSLMTQIDNQLQADPLAEIRTSITNVSENIOTVNVKLLKNOQSAALQOIQKVOVDNN
AOLMWKIQLOQDGLRXYTAGIGAGIENTPAGQSQVOLLAADEIAMIINPANGNTKPM
FVQGOQIFPNEVFLXYLTAPTTISGNGPPAFSLTDPGLTAKNADISNVNANSXPL
NVTYINENCVRGLKLANQIEGLDVLVTKGFRPRDSRAPERNPSGTTIVRYVDQDFD
RQIVTAPVAFSGAKHEKHTDIYSSCLIVRNKGAETIYNRTALDNTLIYSGVIDMPAG
HGHTLEFSYSAWLNVNWPYASISDLVVVMKKATAGITIS"
/ gene
3661. .4260
/ gene="21917"
3661. .4260
/ gene="21917"
/ function="putative membrane; Other or unknown (Phage or
prophage Related)"
/ notes="Residues 1 to 199 are 75.87 pct identical to
residues 1 to 199 of 199 from GenPept 118 ;
g117532789|gb|AA63231.1|AF151091_2 (AF151091) Lom
[prophage P-E1bA]"
/ codon_start=1
/ transl_table=11
/ product="putative outer membrane protein of prophage
CP-933X"
/ protein_id="AAG56006.1"
/ db_xref="GI:12514846"
/ translation="MRKLCVILSAVVMQVAAATPASAAHQSTLSAGYLHASTNVPK
SDLLGNVRYEPMDFMLGLITSFVANAEDQKRYSDTRWHEDSVNRWFSVWAG
PSVRYNEFSAYAMAGVAYSRSVTFSGDYLRTVNDKGKTHDVLTSDDGRHSNTSLAW
GAGVGNPXSVALIDIAFEGSGDWRDTGFIYGVGYKF"
/ gene
4319. .7240
/ gene="21918"
4319. .7240
/ gene="21918"
/ function="putative membrane; Not classified (Phage or
prophage Related)"
/ notes="Residues 1 to 498 of 973 are 95.78 pct identical to
residues 1 to 498 of 1122 from GenPept 118 ;
g11787636|gb|AAC74454.1| (AE000234) putative membrane
protein [Escherichia coli]"
/ codon_start=1
/ transl_table=11
/ product="putative membrane protein of prophage CP-933X"
/ protein_id="AAG56007.1"
/ db_xref="GI:12514847"
/ translation="MNAVKISGVLDKGTCKPVENCTIOLKARRNSATVVVNTVASEN
PDEAGRYSMDEVGYGVSVILLVEGFPSPHAGTITVYEDSOPGTLNDFLGAMTEDDVRP
EALRFRFLVVEVARNASAYAQNTAAKKKSADASTSAREATHADAASARAATS
AGQASQAQSSAGTASTKATEASKSAASAAESSKSAATSAAGAKTSETNAASV500
SAATSAATTTKASEAASASSADASASKEAAKSETSAASASSASATAGNSAKAA
KTSSTNAKSETAEQASAAAGSKTAAALSAASASTSAGOSASASATAAGKSAESAA
SASATTTKAGEATEQASAAASASAAKTSETNAKASETSAESKTAASASSAAGSA
SSASAKDEATROASAAKSSATTAATAGSATAAQAOSKTAESNATRAETAAR
AEDIASVALEADSTTKGIVOLSSATNSTSESATAPKAVKAYELANKGYTAQDRT
TAQKIVOLSNATNSTSEMLAATPKSVKAAAYDLANGIYTAODATTAQKIVOLSSATN
SASFLTAATPKAVKANDNANGRPVSARKVKSALSDITITPKDGTGLNSTSTFSG
GAGFKLATVTPQAOSVWGLITIGAGFNVGSPQOAGISLVLIRAGNPKGITLAC
WORTSTGTFNFWNTYSGDYDIYVAGTGNVIONDYTNASVTIHTSPVSAK
PEGITDGTVSLTSPSQFPYPPGAPIPWSDTVPVSGYALMGOTFDKSAVPKLAAYP
SGVTPDMRGNTIKGKPSAGVLSQEOQDGIKSHTHSASASSTDLGKTTTSSFDYKKS
TNNTGATHSVSGTAAGNHTHSVTGASAVSQWSQNGSVHKVSAASVNTSARAAGHT
HSVSGTAAASGAHAHTVIGIAGHTHSAIGSHGHTITVNAAGNAENTVKNIAFYIVKL
A"
/ complement(4822. .5571)
/ gene="21919"
/ complement(4822. .5571)
/ gene="21919"
/ function="orf; Other or unknown (Phage or Prophage
Related)"
/ notes="Residues 45 to 248 of 249 are 64.21 pct identical
to residues 5 to 205 of 206 from GenPept 118 ;
g11508991|gb|AA96556.1| (J02459) orf206b [bacteriophage
lambda]"
/ codon_start=1

```

```

/transl_table=11
/product="unknown protein encoded by prophage CP-933X"
/protein_id="AAG56008.1"
/db_xref="GI:12514848"
/translation="MDAVVALLFMAADACLIVASSFEADADDAAELADDAADAAVFEDS
ALVSDPAFVSDVFAAEALAAALACSVASPAFVVAVEADDAALSDPAFVALAEAC
PALVDAALADNAAAFEPAPAAALCSAAVSEDLAFVSDVLAFAELPAFVAABEAALLA
LGAALVSDDFAAAFSEADASLAENASDAFVVAVDAEVAACDCCDTAAFVSDVFAAPALV
AAALFSDSAAAALFEDASVAFVDAVPALEADAD"
/ gene
7240. .7821
/ gene="21920"
7240. .7821
/ gene="21920"
/ function="orf; Other or unknown (Phage or Prophage
Related)"
/ notes="Residues 1 to 191 of 193 are 94.76 pct identical to
residues 1 to 191 of 191 from GenPept 118 ;
g112367120|gb|AAC74619.1| (AE000252) orf, hypothetical
protein [Escherichia coli]"
/ codon_start=1
/ transl_table=11
/ product="putative tail fiber protein of prophage CP-933X"
/ protein_id="AAG56009.1"
/ db_xref="GI:12514849"
/ translation="MAFRMSEOPRTIKIYNLLTGTNEFIGEGDAYIPPHGTGLPANSTY
IAPDPDPAGVAVVNDSESWHLVEDHREKTVVDVAGDALFISELGSPLPNTWLSLP
EGEFKQNTAVKDAEAKLFRIAREEETKNSLMOVASEHIAPLQDAVDLEIATEEE
ASLLEAKRYVLLNRVDTSTAPDIEMPIEV"
/ complement(7941. .8831)
/ gene="21921"
/ complement(7941. .8831)
/ gene="21921"
/ function="orf; Other or unknown (Phage or Prophage
Related)"
/ notes="Residues 1 to 289 of 296 are 84.42 pct identical to
residues 1 to 289 of 292 from GenPept 118 ;
g117162108|emb|CAB76676.1| (Y19180) putative catalase
[Salmonella typhimurium]"
/ codon_start=1
/ transl_table=11
/ product="unknown protein encoded by prophage CP-933X"
/ protein_id="AAG56010.1"

```

## Alignment Scores:

Pred. No.:	0.416	Length:	10432
Score:	90.00	Matches:	29
Percent Similarity:	45.92%	Conservative:	16
Best Local Similarity:	29.59%	Mismatches:	52
Query Match:	11.58%	Indels:	1
DB:	1	Gaps:	0

US-09-847-539A-6 (1-159) x AE005333 (1-10432)

Qy	50	AlaThrThrAlaLeGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr	69
Db	5375	GCACGCGATTCTGCTCCGACGCGAAGCATCCGAAACGACGCGTTCGGAAC	5434
Qy	70	AspAlaLeuGlnSerGluGluAlaAlaValLysAlaAspAsnAlaAlaSerAspAla	89
Db	5435	AGCCGAGAATCCTCAAAACCGGTCGCGCATCGTCAGCCAGTTCGCGCGCTCATCGCA	5494
Qy	90	LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln	109
Db	5495	TCATCTCGCTGCTGCTCAAAAGATGAGCGACACAGCGTCCAGCAGCGAAGCAGC	5554
Qy	110	SerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspVal	129
Db	5455	GCCACGACGGCATCCACAGGCGACAGCAGCAGCTGTAGTCGACGACGAGCTCAG	5614
Qy	130	LysLysThrLysAspThrLysProValLysLysGlu-GluArgGlnAsn	146
Db	5615	AGCAAAAGTACGCGGGAATCTGCAGACACGCGCGCTGAGACGCGGCAAAAC	5666









```

|||||
Db 1416 GCAACTCTCTCTACAACTGGTGAAGGAAGCAACCCA 1451
|||||
RESULT 24
SG148IGG
LOCUS
DEFINITION SG148IGG 1469 bp DNA linear BCT 12-JUL-1995
ACCESSION X04015
VERSION X04015
KEYWORDS gamma-immunoglobulin binding protein G.
SOURCE Streptococcus sp.
ORGANISM Streptococcus sp.
REFERENCE
AUTHORS Guss,B., Eliasson,M., Olsson,A., Uhlen,M., Frej,A.K., Jorvall,H.,
Flock,J.I. and Lindberg,M.
TITLE Structure of the IgG-binding regions of streptococcal protein G
JOURNAL EMBO J. 5 (7), 1567-1575 (1986)
MEDLINE
COMMENT Repetitive structures of 24 amino acids, A1, A2 and A3, are found
interrupted by two unrelated repeats of 51 amino acids, B1 and B2.
Further downstream, following a 12-residue unique 'spacer'-region'
(S), there is a stretch of 55 amino acids repeated three times, C1,
C2 and C3. The latter regions are interrupted by two 15-residue
regions, called D1 and D2. Directly following these repeats there
is a region W, which consists of a repetitive, extremely
hydrophilic, structure containing several proline residues. This
region resembles a similar structure in the streptococcal M protein
and in protein A, which is thought to mediate the binding of the
protein to the cell wall.
FEATURES
source
1..1469
/organism="Streptococcus sp."
/strain="G 148"
/db_xref="taxon:1306"
<1..1443
/codon_start=1
/transl_table=11
/product="protein G"
/protein_id="CAA27638.1"
/db_xref="GI:47072"
/db_xref="SWISS-PROT:PI1909"
/translation="EFNKYGVSDDYKNNLNNAKTVEGVKDLQAOVVESEAKKARISEAT
DGLSDFLKSTQPAEDTVKSIELAEAKVLANRELDKYGVSDDYKNNLNNAKTVEGVKDL
QAOVVESEAKKARISEATDGLSDFLKSTQPAEDTVKSIELAEAKVLANRELDKYGVSDDY
YKNNLNNAKTVEGVKALIDEILAAALPKTDYKILNGKTLKGETTTEAVDAATAEKVF
KOYANDGVGENTYDDATKTFTVTEKPEVIDASELTPAVTYYKLINGKTLKGETT
EAVDAATAEKVFKOYANDGVGENTYDDATKTFTVTEKPEVIDASELTPAVTYYKL
INGKTLKGETTKAVDAETAERAKFOYANDGVGENTYDDATKTFTVTEKPEVIDASELTP
APTEPEKPEASIPLVPLTPATPIAKDDAKDDTKKEDAKKPEAKKEDAKKAETLPTTG
EGSNPFFTAALAVMAGALAVASRRKED"
BASE COUNT 546 a 279 c 297 g 347 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0504 Length: 1469
Score: 90.00 Matches: 18
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 12
Query Match: 11.58% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6 (1-159) x SG148IGG (1-1469)
Qy 128 AspValLysLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
||| |||||
Db 1270 GATACTAAGAAGAAGATGCTAAAGAACCAAGAGCTAAGAAAGACGCTAAGAAAGCT 1329
||| |||||
Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
||| |||||
Db 1330 GAAACTCTCTCTACAACTGGTGAAGGAAGCAACCCA 1365
||| |||||
Alignment Scores:
Pred. No.: 0.0504 Length: 1469
Score: 90.00 Matches: 18
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 12
Query Match: 11.58% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6 (1-159) x SG148IGG (1-1469)
Qy 128 AspValLysLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
||| |||||
Db 1270 GATACTAAGAAGAAGATGCTAAAGAACCAAGAGCTAAGAAAGACGCTAAGAAAGCT 1329
||| |||||
Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
||| |||||
Db 1330 GAAACTCTCTCTACAACTGGTGAAGGAAGCAACCCA 1365
||| |||||
RESULT 25
AL12446
LOCUS
DEFINITION AL12446 1469 bp DNA linear PAT 04-JAN-1994
ACCESSION AL12446
VERSION AL12446.1
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial construct.
REFERENCE
AUTHORS 1 (bases 1 to 1469)
TITLE METHOD AND MEANS FOR PRODUCING A PROTEIN HAVING THE SAME IgG
SPECIFICITY AS PROTEIN G
Patent: WO 8705631-A 4 24-SEP-1987;
FEATURES
source
1..1469
/organism="synthetic construct"
/db_xref="taxon:32630"
1..1443
/notes="Protein sequence is in conflict with the conceptual
translation"
/transl_table=11
/product="protein having the same IgG specificity as
protein G from group G streptococci strain G148"
/protein_id="CAA01034.1"
/db_xref="GI:4526638"
/translation="EFNKYGVSDDYKNNLNNAKTVEGVKDLQAOVVESEAKKARISEAT
DGLSDFLKSTQPAEDTVKSIELAEAKVLANRELDKYGVSDDYKNNLNNAKTVEGVKDL
QAOVVESEAKKARISEATDGLSDFLKSTQPAEDTVKSIELAEAKVLANRELDKYGVSDDY
YKNNLNNAKTVEGVKALIDEILAAALPKTDYKILNGKTLKGETTTEAVDAATAEKVF
KOYANDGVGENTYDDATKTFTVTEKPEVIDASELTPAVTYYKLINGKTLKGETT
EAVDAATAEKVFKOYANDGVGENTYDDATKTFTVTEKPEVIDASELTPAVTYYKL
INGKTLKGETTKAVDAETAERAKFOYANDGVGENTYDDATKTFTVTEKPEVIDASELTP
APTEPEKPEASIPLVPLTPATPIAKDDAKDDTKKEDAKKPEAKKEDAKKAETLPTTG
EGSNPFFTAALAVMAGALAVASRRKED"
BASE COUNT 545 a 280 c 297 g 347 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0504 Length: 1469
Score: 90.00 Matches: 18
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 12
Query Match: 11.58% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6 (1-159) x AL12446 (1-1469)
Qy 128 AspValLysLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
||| |||||
Db 1270 GATACTAAGAAGAAGATGCTAAAGAACCAAGAGCTAAGAAAGACGCTAAGAAAGCT 1329
||| |||||
Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
||| |||||
Db 1330 GAAACTCTCTCTACAACTGGTGAAGGAAGCAACCCA 1365
||| |||||
Alignment Scores:
Pred. No.: 0.0504 Length: 1469
Score: 90.00 Matches: 18
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 12
Query Match: 11.58% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6 (1-159) x AL12446 (1-1469)
Qy 128 AspValLysLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
||| |||||
Db 1270 GATACTAAGAAGAAGATGCTAAAGAACCAAGAGCTAAGAAAGACGCTAAGAAAGCT 1329
||| |||||
Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
||| |||||
Db 1330 GAAACTCTCTCTACAACTGGTGAAGGAAGCAACCCA 1365
||| |||||
RESULT 26
AY060997
LOCUS
DEFINITION AY060997 2022 bp mRNA linear INV 08-NOV-2001
ACCESSION AY060997
VERSION AY060997.1
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
REFERENCE
AUTHORS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 2022)
Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,

```







```

[prophage P-Eiba]}
/codon_start=1
/transl_table=11
/product="putative outer membrane protein of prophage
CP-933k"
/feature="Other or unknown (Phase or Prophage
Related)"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKCAAILSAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
SDLLNGINVKYIYETDILGLITSESYANAEDQTHYSDTRHEDYVNRNRFVSVAG
PSVRVNEFSATAMAGVAYSRVSTFGDFRVTDNKRKTHDVLTGSDDARYSNTSLAW
GAGVQNFTEVSAYDVAVEXSGSDMRDGTGTVGVKPF"
719..2041
/gene="20982"
CDS
719..2041
/function="putative structure; Structural component (Phase
or Prophage Related)"
/feature="20982"
/feature="putative structure; Structural component (Phase
or Prophage Related)"
/feature="Residues 164 to 440 are 68.79 pct identical
to residues 381 to 645 from GenPept 118 ;
g114585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933k"
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAVQISGLVKDAGKPIQNCITQLKARNSTXVVNTVASE
NDEAGRYSMDVEYGOYSVTLIVGFPFPHAGTISVYEDSQPTLNDFIGAMTEDDAR
PEALRFQOMVEEAEHAEAEKKNAGEAETSARNAGISAKSAEASANAADTSAEDAISE
SARQAESAASAEASASSSSASAAKASBSLOSATDAELSKTAEASAGNAARDA
TWTSTKARSASASAEOSRTAAEDAVNRITPVVPGPKGPGAGPKGPKGDKGE
RQDTPAGATGRRGQDTPAGPKGPKGDRGERTGLTGAGPKGPKGDKTGAAGPA
GQGPKGERTGAAGVAGTGPQPKGPKGSETQTRFRGLPMRIETISYGFPGTGNALI
TGLTLDPADATQVQOMFQHLQVRFGDGPQVQKGLDEVGSDTGRTE"
2004..2312
/gene="20984"
CDS
2004..2312
/feature="20984"
/function="Other or unknown (Phase or Prophage
Related)"
/feature="Residues 1 to 102 of 102 are 98.03 pct identical
to residues 28 to 129 from GenPept 118 ;
g114585437|gb|AAD25465.1|AF125520_60 (AF125520)
hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MKWAVIOAELENDMNILKLMQSLCCGCKHDDCNGRSITQALR
LGPADILESDENGIIPQQRDRVITQVYILDADRKKIQCVVRPLQILRADGTWENIGMK
"
2489..3469
/gene="20985"
CDS
2489..3469
/feature="20985"
/function="Other or unknown (Phase or Prophage
Related)"
/feature="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from GenPept 118 ;
g11690367|gb|AAF33527.1| (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPIRTFFHNSVINIQQSPQCVTSFACKVELKVIDEKTPILF
QNFENPERKKEDEPIVNTKQHPYLDNVTNAARESDRMIGIFVGDGFSVOKTAFS
KLERFENMIIYREDVDFSMYDRKLSDIYHD1ICEORLRTEDKREDYLLNLEKELR
EISKAQDSLSIYAKRNHAWDFFRNALAKGAEFRCTYNTKNHGISFGGCTYLD
MDMLITGLGTIYAPDGLISMHVDRNDSVNIENSALIVNRSNHPALLEGLSFMHSKVD
AHPYVDGLGKGVKVFNFETPLNHNHFCDFIEFNHPIIMNTSQYTCSSW"

```

```

gene
CDS
3530..4522
/gene="20986"
3530..4522
/gene="20986"
/function="Other or unknown (Phase or Prophage
Related)"
/feature="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPSLQSNFNSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
SAADLSNVYESVIRAVHDSRSLIDQHTVDMIGNTVLDALRSQTFRDVAVSGIHNKX
VHIGCIKYNEYELNESSVKIDDQSLCTCNELYEDVGOEPIFFICEAGENDNEEPPY
VFSVAPDSDSYEMPSWQBLHEITHVHTSSDPSGDSNIELGPTTEILARRVQELG
WSPVDFKGYAEPERAEHRLRLNALRQAAHRENERAFFERLGTISDRYEAQSPDFT
EYSVSNIGYGFIIQHDFFGLAINDLQDANQIQLYHGAPYITFTFGDVKHNOQ"
5350..6231
/gene="20989"
CDS
5350..6231
/gene="20989"
/function="Other or unknown (Phase or Prophage
Related)"
/feature="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYSVNLGCSWNSLTRNLTPDNRLVSSVRDAAVHSDNGAQQ
KVNRTYRVATDNPDPVYAPPELPSVDVNLRSVPGNVKGVKNAVYMLNADSPYSAM
ERGIHVSNRDLPDPVYAPPELPSVDVNLRSVPGNVKGVKNAVYMLNADSPYSAM
FTTSSNEEVTSEVRCFNOYAGSAEKIYGNNGDILGIRMDKINGESLNLISLPPAQ
AEHAIYDMFDRLEQKGLFVDTTETNVLVDRAKNENFIDISSYNNVSDRSWSESQIMQ
SYHGGKQDLISVVLSKI"
6462..7160
/gene="20990"
CDS
6462..7160
/gene="20990"
/function="Other or unknown (Phase or Prophage
Related)"
/feature="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNLVLHQSRSSSMSDFTDIESLVKASSVOMIKNNPOLRF
QGTDNIIYQIEAALDKIGSTETGRVLLNAIESTSRKSETVVIHLNLSRLGVMAIHRD
IDAEHNRGTGSDPHCNLNAVEYPCGEGISVDPFHATIVFHELLHFNHNGERLKVES
SRAESQYSPILLLEARTVGLGAFSEVLENKFHIEIGMPRRTSYPPXDSALIHDDNT
VSLGFOQVRLHPLL"
complement(7668..8144)
/gene="ybhb"
/feature="ybhb"
/feature="Unknown function"
/feature="Residues 1 to 158 of 158 are 99.36 pct identical
to residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGDKLPHRHFVNGMYDGNISPHLADDDVPAGTK
SFVVTYCDPADPTGSGMHWVNVNLPADTRVLPGQFGLVAMPDGLVLTQTRDFGKTG
YDGAAPPKGETHYRFTVHALDVERIDVDDEGASGAVGVNFVHISLASISATAMES"
complement(8203..9492)
/gene="b10A"
/feature="20993"

```













----- Summary Statistics  
 Center project name: H\_NH0461M18  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://daccpac.med.buffalo.edu>)

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-341M1, 2000 bp overlap; the clone sequenced to the right is RP11-289K3. Actual start of this clone is at base position 162181 of RP11-341M1; actual end is at base position 167227 of RP11-461M18.

Data from AC016725 and AC017031 was used to finish this clone, AC020602. There is a tandem repeat from base position 33099 to 33933.

#### FEATURES

source

#### Location/Qualifiers

1. .167227  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"

/clone="RP11-461M18"

/clone\_lib="RPCI-11"

/rpt\_family="MIR"

250. .316

/rpt\_family="MIR"

878. .923

/rpt\_family="MIR"

1376. .1470

/rpt\_family="MIR"

2351. .2391

/rpt\_family="(TTTAA)n"

2850. .3229

/rpt\_family="MaLR"

3260. .3347

/rpt\_family="ERV1"

3488. .3516

/rpt\_family="(TTTG)n"

3490. .3796

/rpt\_family="Alu"

5077. .5106

/rpt\_family="AT\_rich"

5090. .5289

/rpt\_family="L1"

6473. .6782

/rpt\_family="Alu"

6754. .6830

repeat\_region  
 6841. .7120  
 /rpt\_family="GA-rich"  
 repeat\_region  
 7335. .7580  
 /rpt\_family="L1"  
 repeat\_region  
 7792. .7948  
 /rpt\_family="L1"  
 repeat\_region  
 8071. .8364  
 /rpt\_family="Alu"  
 repeat\_region  
 8486. .8631  
 /rpt\_family="L1"  
 repeat\_region  
 8668. .8693  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 8738. .8887  
 /rpt\_family="Alu"  
 repeat\_region  
 8891. .9208  
 /rpt\_family="Alu"  
 repeat\_region  
 9061. .9090  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 9210. .9343  
 /rpt\_family="L1"  
 repeat\_region  
 9649. .9701  
 /rpt\_family="MIR"  
 repeat\_region  
 9702. .9830  
 /rpt\_family="T2\_type"  
 repeat\_region  
 9831. .9992  
 /rpt\_family="MIR"  
 repeat\_region  
 10568. .10588  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 11202. .11499  
 /rpt\_family="Alu"  
 repeat\_region  
 11531. .11609  
 /rpt\_family="Alu"  
 repeat\_region  
 11612. .11641  
 /rpt\_family="(TTTG)n"  
 repeat\_region  
 11619. .11792  
 /rpt\_family="Alu"  
 repeat\_region  
 11793. .11850  
 /rpt\_family="Alu"  
 repeat\_region  
 12082. .12261  
 /rpt\_family="MER1\_type"  
 repeat\_region  
 12485. .12870  
 /rpt\_family="L1"  
 repeat\_region  
 12890. .12924  
 /rpt\_family="(TTTG)n"  
 repeat\_region  
 12921. .12982  
 /rpt\_family="L1"  
 repeat\_region  
 12983. .13285  
 /rpt\_family="Alu"  
 repeat\_region  
 13257. .13292  
 /rpt\_family="(GAA)n"  
 repeat\_region  
 13286. .13552  
 /rpt\_family="L1"  
 repeat\_region  
 13726. .14016  
 /rpt\_family="L1"  
 repeat\_region  
 14020. .14331  
 /rpt\_family="Alu"  
 repeat\_region  
 14335. .14439  
 /rpt\_family="Alu"  
 repeat\_region  
 14509. .15557  
 /rpt\_family="L1"  
 repeat\_region  
 15558. .15850  
 /rpt\_family="Alu"  
 repeat\_region  
 15851. .15904  
 /rpt\_family="L1"  
 repeat\_region  
 15914. .16150  
 /rpt\_family="Alu"  
 repeat\_region  
 16156. .16302  
 /rpt\_family="L1"  
 repeat\_region  
 16313. .16664  
 /rpt\_family="L1"  
 repeat\_region  
 16682. .16939  
 /rpt\_family="L1"

Alignment Scores:

Pred. No.: 1.13e-08 Length: 1576  
Score: 142.00 Matches: 37  
Percent Similarity: 60.61% Conservatives: 3  
Best Local Similarity: 56.06% Mismatches: 26  
Query Match: 18.28% Indels: 1  
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x SG148PG (1-1576)

Qy 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
|||||  
Db 391 GTTGACTCACCAGATACCCCAATTATTCGTAATGGTGTGAATTAACATAATCTT 450  
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleaspGluLeu 40  
|||||  
Db 451 CTGGGAATTCAGACACACACTGCTTGGCGTAATGAGAGAGCTGTACAGCTGATTG 510  
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60  
|||||  
Db 511 ACAGCAGCAGCGGTAGCGGACTGTGGCAGCAGCGGCGAGTGAAATGCTGGGCAGCA- 569  
Qy 61 AlaLeuGluAlaLeuAla 66  
|||||  
Db 570 GCTTGGGAAGCAGCGGCA 587

RESULT 14

STRABP 969 bp DNA linear BCT 26-APR-1993  
LOCUS Streptococcus canis (group G) albumin-binding protein gene, partial cds.  
DEFINITION M95520.1 GI:153554  
ACCESSION albumin-binding protein.  
VERSION Streptococcus canis (group G) (individual isolate DG12) DNA.  
KEYWORDS Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
SOURCE Streptococcus.  
ORGANISM Streptococcus.  
REFERENCE 1 (bases 1 to 969)  
AUTHORS Sjobring,U.  
TITLE Isolation and molecular characterization of a novel albumin-binding protein from group G streptococci  
JOURNAL Infect. Immun. 60, 3601-3608 (1992)  
MEDLINE 92363555  
FEATURES Location/Qualifiers  
source 1..969  
/organism="Streptococcus canis"  
/isolate="DG12"  
/db\_xref="taxon:1329"  
cDS <1..969  
/codon\_start=1  
/transl\_table=1  
/product="albumin-binding protein"  
/protein\_id="AAA26847.1"  
/db\_xref="GI:153554"

CDS

SVKIKIASVDREITELSAQDQIVSAQDNEAITRAEDSSKAAADQANTAKA  
EADLAKARESSDAWEKAAALDQAKAALKEFDYGVSNYKNIINKAKTVEGIMEL  
QAQVVEAKKARISEATDGLSDFLKSQTSADTLKSIKLSAKEMAIRELDAQVSDF  
YKNNKNTVEGVVLDLNLISLAKPEVPEKPEVPEKPEVPEVPEVPEVPEVPEV  
KPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEV  
BASE COUNT 359 a 168 c 219 g 223 t  
ORIGIN

Alignment Scores:

Pred. No.: 7.62e-07 Length: 969  
Score: 126.00 Matches: 26  
Percent Similarity: 65.57% Conservatives: 14  
Best Local Similarity: 42.62% Mismatches: 21  
Query Match: 16.22% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x STRABP (1-969)

Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 81  
|||||  
Db 184 TTGTGCGCGCAGCAGACCATTTCTTCACCAAGCAGATAACAGAGCTATTACTAAG 243  
Qy 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101  
|||||  
Db 244 GCAGAAGAAGATTCTTCCCAAGGCTGGGAAGCTGCTGCGGATCAAGCAAAATACTGCTAAA 303  
Qy 102 SerGluGluAlaGluValGlnSerAspAsnAlaAlaSerAspAlaThrGlnLysAla 121  
|||||  
Db 304 GCAGAAGCAGATGAACCTTGCAAAGGCGAGAAAAGAATCATCAGATGCTTGGGAGAAGGCT 363  
Qy 122 Ala 122  
|||||  
Db 364 GCT 366

RESULT 15

AC020602/c 167227 bp DNA linear PRI 10-JAN-2002  
LOCUS Homo sapiens BAC clone RP11-461M18 from 2, complete sequence.  
DEFINITION AC020602  
ACCESSION AC020602.6 GI:17386508  
VERSION HTG.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 167227)  
AUTHORS Sulston,J.E. and Waterston,R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
REFERENCE 2 (bases 1 to 167227)  
AUTHORS Walligorski,J., Cotton,M. and Doeber,A.  
TITLE The sequence of Homo sapiens BAC clone RP11-461M18  
JOURNAL Unpublished (2002)  
REFERENCE 3 (bases 1 to 167227)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (05-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 167227)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (06-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 5 (bases 1 to 167227)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (07-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 6 (bases 1 to 167227)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 7 (bases 1 to 167227)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (10-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT On Dec 6, 2001 this sequence version replaced gi:14029092.

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)

```

repeat_region 1252..1416
                /note="repetitious region 3"
repeat_region 1417..1461
                /note="repetitious region 4"
repeat_region 1462..1626
                /note="repetitious region 3"
repeat_region 1627..1671
                /note="repetitious region 4"
repeat_region 1672..1836
                /note="repetitious region 3"
misc_feature 1837..2046
                /note="cell-wall spanning region"
misc_feature 2047..2124
                /note="membrane spanning region"
repeat_unit 2157..2165
                /note="inverted repeat A"
repeat_unit 2172..2180
                /note="inverted repeat A"
BASE COUNT 907 a 412 c 491 g 646 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 3,9e-10 Length: 2456
Score: 155.00 Matches: 37
Percent Similarity: 62.12% Conservative: 4
Best Local Similarity: 56.06% Mismatches: 25
Query Match: 19.95% Indels: 0
DB: 1 Gaps: 0

```

US-09-847-539A-6 (1-159) x SGPRTG (1-2456)

```

QY 1 ValAspSerProtleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
    |||||
Db 445 GTTGACTCACCACATCGAAGATACCCCAATATTTCGTAATGCTGGTGAATTAATCTT 504
    |||||

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaLeuAspGluLeu 40
    |||||
Db 505 CTGGGCAATTCAGAGACACACTGGCTTTCGTAATCAAGAGAGTGTACAGCTGATTG 564
    |||||

QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaLeuSerSerAsp 60
    |||||
Db 565 ACAGCAGCAGCGGTGATGCTGGCAGCAGCGCAGCTGAAAATGCTGGGCGAGCA 624
    |||||

QY 61 AlaLeuGluAlaLeuAla 66
    |||||
Db 625 GCTTGGGAAGCAGCGGCA 642
    |||||

```

```

RESULT 12
I08536
LOCUS I08536 1950 bp linear PAT 02-DEC-1994
DEFINITION Sequence 2 from Patent WO 8705025.
ACCESSION I08536
VERSION I08536.1 GI:588754
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1950)
AUTHORS Farnestock,S.R.
TITLE CLONED STREPTOCOCCAL GENES ENCODING PROTEIN G AND THEIR USE TO
CONSTRUCT RECOMBINANT MICROORGANISMS TO PRODUCE PROTEIN G
JOURNAL Patent: WO 8705025-A 2 27-AUG-1987;
FEATURES
    source
        location/Qualifiers
            1..1950
                /organism="unknown"
BASE COUNT 706 a 324 c 397 g 523 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 5.5e-10 Length: 1950
Score: 153.00 Matches: 37
Percent Similarity: 62.12% Conservative: 4
Best Local Similarity: 56.06% Mismatches: 25

```

```

Query Match: 19.69% Indels: 0
DB: 6 Caps: 0
US-09-847-539A-6 (1-159) x I08536 (1-1950)
QY 1 ValAspSerProtleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
    |||||
Db 677 GTTGACTCACCACATCGAAGATACCCCAATATTTCGTAATGCTGGTGAATTAATCTT 736
    |||||

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaLeuAspGluLeu 40
    |||||
Db 737 CTGGGCAATTCAGAGACACACTGGCTTTCGTAATCAAGAGAGTGTACAGCTGATTG 796
    |||||

QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaLeuSerSerAsp 60
    |||||
Db 797 ACAGCAGCAGCGGTGATGCTGGCAGCAGCGCAGCTGAAAATGCTGGGCGAGCA 856
    |||||

QY 61 AlaLeuGluAlaLeuAla 66
    |||||
Db 857 GCTTGGGAAGCAGCGGCA 874
    |||||

RESULT 13
SGI48PG
LOCUS SGI48PG 1576 bp DNA linear BCT 23-MAR-1993
DEFINITION Streptococcus G148 protein G' structural gene.
ACCESSION X53324
VERSION X53324.1 GI:288358
KEYWORDS Streptococcal protein G.
SOURCE Streptococcus sp. 'group G'.
ORGANISM Streptococcus sp. 'group G'.
REFERENCE 1 (bases 1 to 1576)
AUTHORS Goward,C.R., Murphy,J.P., Atkinson,T. and Barstow,D.A.
TITLE Expression and purification of a truncated recombinant
        Streptococcal protein G
JOURNAL Biochem. J. 267 (1), 171-177 (1990)
MEDLINE 90226312
FEATURES
    source
        location/Qualifiers
            1..1576
                /organism="Streptococcus sp. 'group G'"
                /strain="G148"
                /db_xref="taxon:1320"
                /clone="SPG3"
                252..1262
                /gene="Protein G'gene"
                292..603
                /gene="Protein G'gene"
                /note="Truncated gene"
                /codon_start=1
                /transl_table=11
                /product="Protein G'"
                /protein_id="CAA37409.1"
                /db_xref="GI:288359"
                /db_xref="SPTREMBL:Q54180"
                /translation="MEKEKKVYFLKKSAFGLASVSAAFVGVSTVFVAVDSPIEDTPII
                L"
                /translation="MEKEKKVYFLKKSAFGLASVSAAFVGVSTVFVAVDSPIEDTPII
                L"
                /gene="Protein G'gene"
                /note="Truncated gene; Start codon TTG not ATG"
                /codon_start=1
                /transl_table=11
                /product="Protein G'"
                /protein_id="CAA37410.1"
                /db_xref="GI:581662"
                /db_xref="SPTREMBL:Q54181"
                /translation="MKGETTTEAVDAATAEKVFKQYANDNGVDGTTYDDATKTFVT
                EKPEVIDASELTPAVTYYKLIVNGTKLKGETTEAVDAATAEKVFKQYANDNGVDGEW
                TYDDATKTFVTKEPEVIDASELTPAVTYYKLIVNGTKLKGETTEAVDAATAEKVFKQYANDNGVDGEW
                QYANDNGVDGVTYDDATKTFVTE"
BASE COUNT 572 a 273 c 327 g 404 t
ORIGIN

```

repeat\_region 1111..1221  
/note="A2 repeated sequence"  
repeat\_region 1354..1446  
/note="A3 repeated sequence"  
repeat\_region 1483..1647  
/note="B1 repeated sequence"  
repeat\_region 1499..1918  
/note="210 bp duplication"  
/rpt\_unit=1708..1918  
repeat\_region 1693..1857  
/note="B2 repeated sequence"  
repeat\_region 1903..2067  
/note="B3 repeated sequence"  
BASE COUNT 868 a 408 c 482 g 626 t  
ORIGIN

Alignment Scores:  
Pred. No.: 3,78e-10 Length: 2384  
Score: 155.00 Matches: 37  
Percent Similarity: 62.12% Conservative: 4  
Best Local Similarity: 56.06% Mismatches: 25  
Query Match: 19.95% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x SGSPG (1-2384)

QY 1 ValAspSerProIleGluInProArgIleIleProAsnGlyGlyThrLeuThrasnLeu 20  
|||||  
Db 676 GTTGACTCACCATCGAAGATACCCCAATATTTCGTAATGGTGGTGAATTAACATCTT 735  
|||||  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
|||||  
Db 736 CTGGGAATTCAGACAAACACTGGCTTTCGTAATGAAGAGAGTCTACAGCTGATTG 795  
|||||  
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
|||||  
Db 796 ACAGCAGCAGCGGTAGCCGATCTGTGGCAGCAGCGGAGTGAATAATGCTGGGCGAGCA 855  
|||||  
QY 61 AlaLeuGluAlaLeuAla 66  
|||  
Db 856 GCTTGGGAAGCAGCGGCA 873  
|||

RESULT 10  
LOCUS 108537  
DEFINITION Sequence 4 from Patent WO 8705025.  
ACCESSION 108537  
VERSION 108537.1 GI:588755  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2384)  
AUTHORS Fahnestock,S.R.  
TITLE CLONED STREPTOCOCCAL GENES ENCODING PROTEIN G AND THEIR USE TO  
CONSTRUCT RECOMBINANT MICROORGANISMS TO PRODUCE PROTEIN G  
JOURNAL Patent: WO 8705025-A 4 27-AUG-1987;  
FEATURES Location/Qualifiers  
source 1..2384  
/organism="unknown"  
BASE COUNT 868 a 410 c 480 g 626 t  
ORIGIN

Alignment Scores:  
Pred. No.: 3,78e-10 Length: 2384  
Score: 155.00 Matches: 37  
Percent Similarity: 62.12% Conservative: 4  
Best Local Similarity: 56.06% Mismatches: 25  
Query Match: 19.95% Indels: 0  
DB: 6 Gaps: 0

US-09-847-539A-6 (1-159) x 108537 (1-2384)

QY 1 ValAspSerProIleGluInProArgIleIleProAsnGlyGlyThrLeuThrasnLeu 20  
|||||  
Db 676 GTTGACTCACCATCGAAGATACCCCAATATTTCGTAATGGTGGTGAATTAACATCTT 735  
|||||  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
|||||  
Db 736 CTGGGAATTCAGACAAACACTGGCTTTCGTAATGAAGAGAGTCTACAGCTGATTG 795  
|||||  
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
|||||  
Db 796 ACAGCAGCAGCGGTAGCCGATCTGTGGCAGCAGCGGAGTGAATAATGCTGGGCGAGCA 855  
|||||  
QY 61 AlaLeuGluAlaLeuAla 66  
|||  
Db 856 GCTTGGGAAGCAGCGGCA 873  
|||

RESULT 11  
SGPROTG  
LOCUS 2456 bp DNA linear BCT 21-MAR-1995  
DEFINITION Streptococcus G148 gene for protein G.  
ACCESSION X06173 M28331  
VERSION X06173.1 GI:47084  
KEYWORDS G protein; IgG receptor.  
SOURCE Streptococcus sp. G148.  
ORGANISM Streptococcus sp. G148.  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 2456)  
AUTHORS Olsson,A., Eliasson,M., Guss,B., Nilsson,B., Hellman,U.,  
Lindberg,M. and Uhlen,M.  
TITLE Structure and evolution of the repetitive gene encoding  
streptococcal protein G  
JOURNAL Eur. J. Biochem. 168 (2), 319-324 (1987)  
MEDLINE 88029445  
FEATURES Location/Qualifiers  
source 1..2456  
/organism="Streptococcus sp. G148"  
/strain="G148"  
/db\_xref="taxon:1324"  
/clone="pSPG1 and pSPG8"  
335..340  
/note="pot. ribosome binding site"  
346..444  
346..2127  
/code="G preprotein (AA -33 to 560)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA29540.1"  
/db\_xref="GI:47085"  
/db\_xref="SWISS-PROT:P19909"  
/translation="MEKKKVKYFLRKSAFGLASVSAAFLVGSTVFAVDSPIEDTPII  
RNGELTNLGNSETTLALRNESATADITAAAVDTAAATAAANAAGAAWAAAAAD  
ALAKAKADALKEPNKYGSDYKLNINNAKTVEGYKDLQAQVVEAKKARISEATDGL  
SDFLSQTPAEDTVKSLAEAKVLANRELDYGVSDYKLNINNAKTVEGYKDLQAQ  
VVEAKKARISEATDGLSDFLSQTPAEDTVKSLAEAKVLANRELDYGVSDYKLN  
LINNAKTVEGYKDLSEILALPQTDYKLLINGKTLKGETTTEAVDAATBKVFQY  
ANDNGVDGWTYDDATKTFTVTEKPEVIDASELTAVTTPVTKLVINGKTLKGETT  
DAATAEKVQYANDNGVDGWTYDDATKTFTVTEKPEVIDASELTAVTTPVTKLVING  
KTLKGETTTEAVDAATAEKVQYANDNGVDGWTYDDATKTFTVTEKPEVIDASELT  
EPEKPEASIPVLPPLTPATPIAKDDTKKEDAKKPEAKKEDAKKAEATLPTTGEGS  
445..2124  
/product="mature G protein (AA 1 - 560)"  
694..765  
/note="repetitious region 1"  
766..918  
/note="repetitious region 2"  
919..990  
/note="repetitious region 1"  
991..1143  
/note="repetitious region 2"  
1144..1215  
/note="repetitious region 1"

RBS  
sig\_peptide  
CDS  
mat\_peptide  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region

```
/protein_id="AAA03664.1"
/db_xref="GI:153823"
translation="MEKEKKVYFLRKSAFGLASVSAFLVGVTFVAVDSPIEDTPII
RNGGELNLLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGAAWAAAAAD
ALAKAKADALKEFNKYGVSDYKNNLNNAKTVEGKDLQAQVVSASAKKARISEATDGL
SDFLSQTPAEDTVKSIELAEAKVLAMRELDKYGSDYHKNNLNNAKTVEGKDLQAO
ILAAALPTDYYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEMTYDDATK
TFTVEKPEVIDASELTPAVTTYKLIVNGKTLKGETTTKAVDAETAASKAKFOYANDNG
VDGVMYDDATKTFTVTEMVTEVPGDAPTPEKPEASTPLVPLTPATPLAKDDAKKD
TKEDAKPEAKKDDAKKAETLPTTGSGSNPFFTAALAVMAGALAVASKRKE"
mat_peptide
BASE COUNT 706 a 323 c 398 g 523 t
ORIGIN 1 bp upstream of HindIII site.
```

```
Alignment Scores:
Pred. No.: 3,04e-10 Length: 1950
Score: 155.00 Matches: 37
Percent Similarity: 62.12% Conservative: 4
Best Local Similarity: 56.06% Mismatches: 25
Query Match: 19.95% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x STRSPGIGP (1-1950)
Qy 1 ValAspSerProileGluGlnProArgileileProAsnGlyGlyThrLeuThrAsnLeu 20
Db 677 GTTGATTACCAATCGAGATACCCCAATATTTCGTAAATGGTGGAATTAATCTT 736
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Db 737 CTGGGAATTCAGACACACACTGGCTTTGCGTAATGAAGAGAGTGCTACAGCTGATTG 796
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaSerSerAsp 60
Db 797 ACAGCAGCAGCGGTAGCCGACTACTGTGCAGCAGCGGAGCTGAAATGCTGGGCAGCA 856
Qy 61 AlaLeuGluAlaLeuAla 66
Db 857 GCTTGGGAAGCAGCGGCA 874
```

```
RESULT 8
LOCUS I09115 1950 bp linear PAT 02-DEC-1994
DEFINITION Sequence 23 from Patent WO 8810306.
ACCESSION I09115
VERSION I09115.1 GI:588187
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1950)
AUTHORS Fahnstock,S.R.
TITLE CLONED PROTEIN G VARIANT GENES AND THE PROTEIN G VARIANTS EXPRESSED THEREFROM
JOURNAL Patent: WO 8810306-A 23 29-DEC-1988;
FEATURES Location/Qualifiers
source 1..1950
/organism="unknown"
BASE COUNT 706 a 323 c 398 g 523 t
ORIGIN
```

```
Alignment Scores:
Pred. No.: 3,04e-10 Length: 1950
Score: 155.00 Matches: 37
Percent Similarity: 62.12% Conservative: 4
Best Local Similarity: 56.06% Mismatches: 25
Query Match: 19.95% Indels: 0
DB: 6 Gaps: 0

US-09-847-539A-6 (1-159) x I09115 (1-1950)
```

```
Qy 1 ValAspSerProileGluGlnProArgileileProAsnGlyGlyThrLeuThrAsnLeu 20
```

```
Db 677 GTTGATTACCAATCGAGATACCCCAATATTTCGTAAATGGTGGAATTAATCTT 736
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Db 737 CTGGGAATTCAGACACACACTGGCTTTGCGTAATGAAGAGAGTGCTACAGCTGATTG 796
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaSerSerAsp 60
Db 797 ACAGCAGCAGCGGTAGCCGACTACTGTGCAGCAGCGGAGCTGAAATGCTGGGCAGCA 856
Qy 61 AlaLeuGluAlaLeuAla 66
Db 857 GCTTGGGAAGCAGCGGCA 874

RESULT 9
LOCUS SGSPG 2384 bp DNA linear BCT 09-OCT-1997
DEFINITION Streptococcus sp. spg gene for protein G.
ACCESSION Y00428
VERSION Y00428.1 GI:47100
KEYWORDS g protein; igg binding protein; spg gene.
SOURCE Streptococcus sp. GX7805.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 2384)
AUTHORS Fahnstock,S.R.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-1987) Fahnstock S.R.; Genex Corp., 16020 Industrial Dr., Gaithersburg, MD 20877, USA
REFERENCE 2 (bases 1 to 2384)
AUTHORS Filipula,D., Alexander,P. and Fahnstock,S.R.
TITLE Nucleotide sequence of the protein G gene from Streptococcus GX7805, and comparison to previously reported sequences
JOURNAL Nucleic Acids Res. 15 (17), 7210 (1987)
FEATURES MEDLINE 88015586
Location/Qualifiers
source 1..2384
/organism="Streptococcus sp. GX7805"
/strain="GX7805"
/db_xref="taxon:1325"
577..2358
/gene="spg"
577..675
/gene="spg"
577..2358
/gene="spg"
/codon_start=1
/transl_table=11
/product="protein G precursor"
/protein_id="CAA68489.1"
/db_xref="GI:47101"
/transl="MEKEKKVYFLRKSAFGLASVSAFLVGVTFVAVDSPIEDTPII
RNGGELNLLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGAAWAAAAAD
ALAKAKADALKEFNKYGVSDYKNNLNNAKTVEGKDLQAQVVSASAKKARISEATDGL
SDFLSQTPAEDTVKSIELAEAKVLAMRELDKYGSDYHKNNLNNAKTVEGKDLQAO
LVESAKKARISEATDGLSDFLSQTPAEDTVKSIELAEAKVLAMRELDKYGSDYKNN
LINNAKTVEGKDLQAO
LVESAKKARISEATDGLSDFLSQTPAEDTVKSIELAEAKVLAMRELDKYGSDYKNN
ANDNGVDGEMTYDDATKTFTVEKPEVIDASELTPAVTTYKLIVNGKTLKGETTTEAV
DAATAEKVFKQYANDNGVDGEMTYDDATKTFTVEKPEVIDASELTPAVTTYKLIVNG
KTLKGETTTEAVDAATAEKVFKQYANDNGVDGEMTYDDATKTFTVEKPEVIDASELTPAV
EPEKPEASTPLVPLTPATPLAKDDAKKDDTKEDAKPEAKKDDAKKAETLPTTGGS
676..2355
/product="protein G"
/gene="spg"
886..996
/feature="Al repeated sequence"
986..1435
/feature="225 bp duplication"
/rpt_unit=1211..1435
mat_peptide
repeat_region
repeat_region
```



```

Db 61 CTTGGCAATGCTCCAGAAAACCTGCAATAGTAAAGAAAGAGCCATTGATCAATTA 120
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaSerSerAsp 60
Db 121 AAAAACAAGCTATTGAGGATAAAGAGCTACGACAGCTATAGAAGCAGCAAGTTTCAGAT 180
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
Db 181 GCCTTAGAAGCATTAGCGGATCAAGCAGCAGCTTTACAATCAGAAGAGCTGCAGTAGTT 240
Qy 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
Db 241 CAATCAGATAACGCTGCTAGTACGCTTTAGAGCATTAGCGGATCAAGCAGCAGCTTTTA 300
Qy 101 GlnSerGluGluAlaGluAlaValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
Db 301 CAATCAGAGAGCTGCGAGTAGTACATCAGATACAGCTGCTGTTGAGCGCTTTAGAAGCA 360
Qy 121 AlaAla 122
Db 361 TTGGCG 366

RESULT 6
AF124402 717 bp DNA linear BCT 14-AUG-2000
LOCUS Streptococcus pyogenes strain KTL9 GRAB precursor, gene, partial
DEFINITION
ACCESSION AF124402
VERSION AF124402.1 GI:4589084
KEYWORDS Streptococcus: pyogenes.
SOURCE Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ORGANISM Streptococcus.
REFERENCE 1 (bases 1 to 717)
AUTHORS Rasmussen, M., Muller, H.P. and Bjorck, L.
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the
JOURNAL bacterial surface by binding alpha2-macroglobulin
MEDLINE J. Biol. Chem. 274 (22), 15336-15344 (1999)
PUBMED 99269061
10336419
REFERENCE 2 (bases 1 to 717)
AUTHORS Rasmussen, M., Muller, H.P. and Bjorck, L.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES
source
location/Qualifiers
/organism="Streptococcus pyogenes"
/strain="KTL9"
/db_xref="taxon:1314"
1..>717
/product="GRAB"
mat_peptide
misc_feature
1..174
/region="Region: alpha2-macroglobulin-binding A domain"
CDS
<1..>717
/note="grab"
/codon_start=1
/transl_table=11
/product="GRAB precursor"
/protein_id="AAD36341.1"
/db_xref="GI:4589085"
1EDKATTAIAASDALEALADQALQSEAAVQSDNAAALADQDQALQS
EEAAVVKADNAASDTLEALADQALQSEAAVQSDNAAADTLEALADQDQALQSE
AAVVKADNAASDTLEALADQALQSEAAVQSDNAAADQALQSEAAVQSDNAAADQALQSE
TKPVVKKEERQNVNTLPTTGE"
misc_feature 175..594
/note="Region: repeat motif"
misc_feature 595..717
/region="Region: cell-wall attachment"
BASE COUNT 257 a 140 c 163 g 157 t

```

```

ORIGIN
Alignment Scores: 2,09e-59 Length: 717
Pred. No.: 534.00 Matches: 112
Score: 534.00 Conservative: 4
Percent Similarity: 95.08% Mismatches: 6
Best Local Similarity: 91.80% Indels: 0
Query Match: 68.73% Gaps: 0
DB: 1

US-09-847-539A-6 (1-159) x AF124402 (1-717)
Qy 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
Db 1 GTTGACTCACCTATCGAAGCAGCTCGAATTATTCGAATGCGGAACCTTAACATACTT 60
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluAlaAlaSerSerAsp 40
Db 61 CTTGGCAATGCTCCAGAAAACCTGCAATAGTAAAGAAAGAGCCATTGATCAATTA 120
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
Db 121 AAAAACAAGCTATTGAGGATAAAGAGCTACGACAGCTATAGAAGCAGCAAGTTTCAGAT 180
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
Db 181 GCCTTAGAAGCATTAGCGGATCAAGCAGCAGCTTTACAATCAGAAGAGCTGCAGTAGTT 240
Qy 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
Db 241 CAATCAGATAACGCTGCTAGTACGCTTTAGAGCATTAGCGGATCAAGCAGCAGCTTTTA 300
Qy 101 GlnSerGluGluAlaGluAlaValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
Db 301 CAATCAGAGAGCTGCGGTTGTTAAAGCGGATACGCTGCTAGTGACACACTTAGAAGCA 360
Qy 121 AlaAla 122
Db 361 TTGGCG 366

RESULT 7
STRSPGIGP 1950 bp DNA linear BCT 23-DEC-1993
LOCUS Streptococcus sp. (Lancefield group G) spg gene encoding an
DEFINITION immunoglobulin G binding protein.
ACCESSION M13825
VERSION M13825.1 GI:153822
KEYWORDS constant region; immunoglobulin binding protein.
SOURCE Streptococcus sp. (Lancefield group G; strain GX7809) DNA, clone
MGX4547.
ORGANISM Streptococcus sp.
Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 1950)
AUTHORS Fahnestock, S.R., Alexander, P., Nagle, J. and Filpula, D.
TITLE Gene for an immunoglobulin-binding protein from a group G
JOURNAL streptococcus
MEDLINE J. Bacteriol. 167, 870-880 (1986)
COMMENT 86304178
Draft entry and computer-readable sequence of [1] kindly provided
by S.R.Fahnestock, 07-NOV-1986. The -35 and -10 regions are
located at positions 465-470 and 487-492 respectively and an SD
sequence at 565-570.
FEATURES
source
location/Qualifiers
/organism="Streptococcus sp."
/db_xref="taxon:1306"
578..576
/note="IgG binding protein signal peptide"
578..1924
/note="IgG binding protein"
/codon_start=1
/transl_table=11
sig_peptide
CDS

```

AUTHORS	Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE	Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin
JOURNAL	J. Biol. Chem. 274 (22), 15336-15344 (1999)
MEDLINE	95269061
PUBMED	10336419
REFERENCE	2 (bases 1 to 469)
AUTHORS	Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE	Direct Submission
JOURNAL	Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegetan 39, Lund 221 00, Sweden
FEATURES	Location/Qualifiers
source	1..469
	/organism="Streptococcus pyogenes"
	/strain="AP1"
	/db_xref="taxon:1314"
mat_peptide	1..>469
misc_feature	1..174
	/product="GRAB"
CDS	<1..>469
	/note="Region: alpha2-macroglobulin-binding A domain"
	/note="grab"
	/codon_start=1
	/transl_table=11
	/product="GRAB precursor"
	/protein_id="AAB26339.1"
	/db_xref="GI:4589081"
	/translation="VDSPTQPIRIIPNGTLLNLGNAPDKALNRNEIRAIDELKKQAEIDKEATTAIEAASDAKALADQALQDSAEAAVYKADNAASDALEALADQDALQSEAEVQSDQNNAAASDAWEAKAATPDAVDRKTKDTKPVYKKERQNVNTLPTGEE"
misc_feature	175..342
	/note="Region: repeat motif"
misc_feature	343..469
	/note="Region: cell-wall attachment"
BASE COUNT	178 a 90 c 100 g 101 t
ORIGIN	
Alignment Scores:	
Pred. No.:	7,16e-88 Length: 469
Score:	754.00 Matches: 155
Percent Similarity:	99.36% Conservative: 0
Best Local Similarity:	99.36% Mismatches: 1
Query Match:	97.04% Indels: 0
DB:	1 Gaps: 0
US-09-847-539A-6 (1-159) x AF124400 (1-469)	
QY	1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
DB	1 GTCGACTCACCTATCGAACAGCCTAGAAATTATTCCAATAATGGCGAACCTTAATTAATCTT 60
QY	21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluArgAlaIleAspGluLeu 40
DB	61 CTTGGCAATGCTCCAGAAAACCTGGCAATTACGTAATCAAGAAGAGAGCATTGATGAATTA 120
QY	41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
DB	121 AAAAAACAAGCTATTGAGGATAGGAAGCTATGCNACGCTATAGAACGACCAAGTTTCAGAT 180
QY	61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIaValVal 80
DB	181 GCCTTTAAGCATTAGCGGATCAACAGACGCTTTCAATCAGAAGAAGCTGCGGTTGTT 240
QY	81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
DB	241 AAACGGGATAACGCTGCTAGTGACGCTTAGAGAGCTTGGCGATCAACAGACGCTTTA 300
QY	101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
DB	301 CAATCACAAGAAGCTGAAGTAGTTCAATCAGATAACGCTCTAGTAGCGCTGGGAAAAA 360
QY	121 AlaAlaThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLys 140

Db	361	CGACCAACTCCAAATCGCTTTTATAGTTTAAAGAAAAACTAAAGATACAAAAACCTGTAGTTAAA	420
Qy	141	LysGluGluAtrGinAsnValAsnThrLeuProThrThrGlyGluGlu	156
Db	421	ANAGAAGAAACAAAAACGTTTAAATACCTTCTTCAACTGGTGAAGAG	468
RESULT 5			
AF124403			
LOCUS			
DEFINITION	AF124403	804 bp DNA linear	BCT 14-AUG-2000
ACCESSION	AF124403	Streptococcus pyogenes strain AP49	GRAB precursor, gene, partial
VERSION	AF124403		
KEYWORDS	AF124403.1	GI:4589086	
SOURCE		Streptococcus pyogenes.	
ORGANISM		Streptococcus pyogenes	
REFERENCE		Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.	
AUTHORS		1 (bases 1 to 804)	
TITLE		Rasmussen, M., Muller, H.P. and Bjorck, L.	
JOURNAL		Protein GRAB of streptococcus pyogenes regulates proteolysis at the	
MEDLINE		bacterial surface by binding alpha2-macroglobulin	
PUBMED		J. Biol. Chem. 274 (22), 15336-15344 (1999),	
REFERENCE		99269061	
AUTHORS		2 (bases 1 to 804)	
TITLE		Rasmussen, M., Muller, H.P. and Bjorck, L.	
JOURNAL		Direct Submission	
FEATURES		Submitted (28-JAN-1999); Cell and Molecular Biology, Molecular	
		Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden	
		Location/Qualifiers	
		1..804	
		/organism="Streptococcus pyogenes"	
		/strain="AP49"	
		/db_xref="taxon:1314"	
		1..>804	
		/product="GRAB"	
		1..174	
		/note="Region: alpha2-macroglobulin-binding A domain"	
		<1..>804	
		/note="grab"	
		/codon_start=1	
		/transl_table=11	
		/product="GRAB precursor"	
		/protein_id="AAB26342.1"	
		/db_xref="GI:458908"	
		/translation="VDSPIEQRIIPNGGTLNLGNAPKIALRNEERAIDELKKQAA	
		IEEKATATEAASDADALADQALQSEAAVYQSDNAASDALEALADQADALQ	
		EEAAVQSDNAGDALEALADQALQSEAAVYKADNAASDALEALADQALQSE	
		ASVVKADNASDALEALADQALQSEAAVYKADNAASDALEALADQALQSEAE	
		VQSDNADNAWEKATPTALDVKKTKDTKPVVKKKEORNVNLTPTTGE"	
		175..678	
		/note="Region: repeat motif"	
		679..804	
		/note="Region: cell-wall attachment"	
		BASE COUNT 284 a 161 c 185 g 174 t	
		ORIGIN	
		Alignment Scores:	
		Pred. No.:	1,31e-59
		Score:	536.00
		Percent Similarity:	94.26%
		Best Local Similarity:	92.62%
		Query Match:	68.98%
		DB:	1
		US-09-847-539A-6 (1-159) x AF124403 (1-804)	
Qy	1	ValAspSerProIleGluGlnProArgIlefileProAsnGlyGlyThrLeuThrAsnLeu	20
Db	1	GTTCACCTACCTATCGAACAGCTTCGAAATATTATTCCAATGCGGAACCTTAACCTT	60
Qy	21	LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu	40

```
Pred. No.: 2,15e-89 Length: 10029
Score: 777.00 Matches: 159
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x AE006573 (1-10029)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
Db 6482 GTTGACTCACCTTTCAGACAGCTCGAATTATTCGAATGGCGAAGCTTAACTAATCTT 6423

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
|||||
Db 6422 CTTGGCAATGCTCCAGAAAGAACTGCGCATACGTAATGAAGAAGAGCCATGATCAATTA 6363

QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
|||||
Db 6362 AAAAACAAGCTATTTCAGGATAAAGAGCTACGACAGCTATAGAAGCAGCAAGTTTCAGAT 6303

QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 80
|||||
Db 6302 GCCTTAGAAGCAATTCAGGATCAACAGACGCTTACAAATCAGAAGAGCTCGGTTGTT 6243

QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
|||||
Db 6242 AAAGCGGATAACGCTAGTCAGCCCTTAGAAGCATTGGCGGATCAACAGACGCTTTA 6183

QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
|||||
Db 6182 CAATCAGAAAGAGCTGAAGTAGTTCATCAGATAACGCTGCTAGTCAGCCCTGGGAAAAA 6123

QY 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140
|||||
Db 6122 GCAGCACTCCAACTGCTTAGATTGTTAGAAAACCTAAAGATACAAAACCTGTAGTTAAA 6063

QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsnPro 159
|||||
Db 6062 AAAGAAGAAAGACAAACGTTAATACCTTCCTACAACTGGTGAAGAGTCTAACCCA 6006

RESULT 3
AF124401 AF124401 468 bp DNA linear BCT 14-AUG-2000
LOCUS Streptococcus pyogenes strain KTL3 GRAB precursor, gene, partial
DEFINITION cds.
ACCESSION AF124401
VERSION AF124401.1 GI:4589082
KEYWORDS Streptococcus pyogenes.
SOURCE Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ORGANISM Streptococcus
REFERENCE 1 (bases 1 to 468)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin
J. Biol. Chem. 274 (22), 15336-15344 (1999)
MEDLINE 99269061
PUBMED 10336419

REFERENCE 2 (bases 1 to 468)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES
source Location/Qualifiers
1..468
/organism="Streptococcus pyogenes"
/strain="KTL3"
/db_xref="taxon:1314"
mat_peptide 1..>468
misc_feature 1..174

CDS
/note="Region: alpha2-macroglobulin-binding A domain"
<1..>468
/codon_start=1
/transl_table=11
/product="GRAB precursor"
/db_xref="GI:4589082"
/translation="VDSPIEQPRIIPNGTITNLGNAPKELALINPERAIDELKKQA
IDEKEATTAIEAASDDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQS
EEAEVQSDNAASDAWEKAATPTALDVKTKDKTPVKVKEERQNVNLTPTGEE"
175..342
/note="Region: repeat motif"
343..468
/note="Region: cell-wall attachment"
BASE COUNT 178 a 91 c 100 g 99 t
ORIGIN
Alignment Scores: 1,21e-88 Length: 468
Pred. No.: 760.00 Matches: 156
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 97.81% Gaps: 0
DB: 1

US-09-847-539A-6 (1-159) x AF124401 (1-468)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
Db 1 GTGGACTCACCTATCGAACAGCCTCGAATTATTCGAATGGCGAAGCTTAACTAATCTT 60

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
|||||
Db 61 CTTGGCAATGCTCCAGAAAGAACTGCGCATACGTAATGAAGAAGAGCCATGATCAATTA 120

QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
|||||
Db 121 AAAAACAAGCTATTTCAGGATAAAGAGCTACGACAGCTATAGAAGCAGCAAGTTTCAGAT 180

QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 80
|||||
Db 181 GCCTTAGAAGCAATTCAGCGGATCAACAGACGCTTACAAATCAGAAGAGCTCGGTTGTT 240

QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
|||||
Db 241 AAAGCGGATAACGCTGCTAGTCAGCGCTTAGAAGCATTGGCGGATCAACAGACGCTTTA 300

QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
|||||
Db 301 CAATCAGAAAGAGCTGAAGTAGTTCATCAGATAACGCTGCTAGTCAGCCCTGGGAAAAA 360

QY 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140
|||||
Db 361 GCAGCACTCCAACTGCTTAGATTGTTAGAAAACCTAAAGATACAAAACCTGTAGTTAAA 420

QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156
|||||
Db 421 AAAGAAGAAAGACAAACGTTAATACCTTCCTACAACTGGTGAAGAG 468

RESULT 4
AF124400 AF124400 469 bp DNA linear BCT 14-AUG-2000
LOCUS Streptococcus pyogenes strain AP1 GRAB precursor, gene, partial
DEFINITION cds.
ACCESSION AF124400
VERSION AF124400.1 GI:4589080
KEYWORDS Streptococcus pyogenes.
SOURCE Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ORGANISM Streptococcus
REFERENCE 1 (bases 1 to 469)
```

```

CDS
complement(619..1911)
/genes="aroA"
/EC_number="2.5.1.19"
/notes="Best Blastp hit = sp|P43905|AROA_LACLA
3-PHOSPHOSHIMKIMATE 1-CARBOXYVINYLTRANSFERASE
(5-ENOLPYRUVYLSHIMKIMATE-3-PHOSPHATE SYNTHASE) (EPSP
SYNTHASE) (EPSP) >gi|1075724|pir|I52580
3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)
- Lactococcus lactis >gi|683583|emb|CAA55180.1| (X78413)
5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus
lactis]"
/codon_start=1
/transl_table=11
/product="putative 3-phosphoshikimate
1-carboxyvinyltransferase"
/protein_id="AAK34180.1"
/db_xref="GI:13622461"
/translation="MKRMKLTNAGPLQGTIQVQDKSISHRAVILGAVAKGETRVKG
LUGEDVLTQAFNLGVRIEEKDQVLIEGGQFQGLNAPCOTLNGNSGTSKRLIA
GLLAGQFVYKMTGDESLSKRPMDRIVYPLKQMGVEISGETDRQFPPLQIQGNRLQP
ITYTLPISSAQVKSAILAALQAKGTQVVEKEITRNHTEEMIQOFGGRLIVDGKRIT
LVGPQOLTAGEITVPCGDISSAEPMLVAGLIPGSELLLNKGVNPTRTGLEVERKG
AQIYVEDMKKEQVTSIRVYVSNMKGTTISGGLIPRLIDELPIALLATOAGCTTICK
DAQLRKVETDRIVQVTDILNSMGANIKATADGMIIGKPTVLYGANTSTYGDHRIQMM
TATAALLVQGVHLDKEAINTSYPTFFKDLERLCHD"
complement(2013..2978)
/genes="SPY1353"
complement(2013..2978)
/notes="SPY1353"
/notes="Best Blastp hit = dbj|BA07074.1| (AP001518)
unknown conserved protein [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34181.1"
/db_xref="GI:13622462"
/translation="MAEKKFDFKLVSKWQVEIQVPMRHLQSAEMLSAIAVAYLIL
TAPPLVIAATFPYPIIDITADLRMKONLKPDIKIPASATIVENIFSKPSGVLGVA
TLTGLTMSRSLTLOKATNKAYGASCHRODFEIGHLVGLTSLILFLLAFALFISF
SKAAJOVLKXHLSDNTITFILLIQPTVLIIFVGLMLLFLLENVKKIKYILP
GLTFSFVTFILSNLVGNVYVNERVDIKMFGVSQVFIIMLWFIPLAKILILGAIF
NATYQMSIGLKLEGRSGDMIALILKLTGNSDLSQSPQSIEDSHTD"
complement(2980..3840)
/genes="map"
/notes="SPY1354"
complement(2980..3840)
/genes="map"
/EC_number="3.4.11.18"
/notes="Best Blastp hit = sp|Q929J4|AMPM_BACHD METHIONINE
AMINOPEPTIDASE (MAP) (PEPTIDASE M)
>gi|11264635|pir|I744405 methionyl aminopeptidase (EC
3.4.11.18) map [imported] - Bacillus halodurans
>gi|4512426|dbj|JBA075293.1| (AB017508) map homologue
(identity of 81% to B. subtilis ) [Bacillus halodurans]
>gi|10172768|dbj|BA03875.1| (AP001507) methionine
aminopeptidase [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative methionine aminopeptidase"
/protein_id="AAK34182.1"
/db_xref="GI:13622463"
/translation="MTLKSAREIEAMDRAQDFLAGIHGLRDLIKPGVDMVEEAYV
RRCKEDNVLPIQIGVDGHMDYPYATCCGLNDEVAHAFPHYIILKEGLLKVDNVLVS
EPLDKSIDVAALDENFENMKWGTSGTGGGLADSCWATVGPSPSEIKKMDVKEA
MYRIGEKAVIGNRIGDIAAQQVEYAESFGYGVVRDLVGHGVPTTHEEPVWPYGTAG
RGLRKEGMLVTEPMINTGTWEIDTIDTKTGAHKTLDGGLSCQYEHQEVITKDGPIV
LTSQSEERTY"
complement(3856..5139)
/genes="SPY1355"
complement(3856..5139)
/genes="SPY1355"
/notes="Best Blastp hit = pir|A69998 hypothetical protein
ytoi - Bacillus subtilis >gi|2293259|gb|AAC00336.1|
(AF008220) YTOI [Bacillus subtilis]
>gi|2635392|emb|CAB14887.1| (Z99118) ytoi [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34183.1"
/db_xref="GI:13622464"
/translation="MSKHODILOYLEKLAIGKVSRSISNHLKVSQDGTAYRAIKRAE
NEGIVETKPRSGTVRIEKKGRVDRILRTYSEIARIISDSSEVLHAGLGHFESRISGA
MTQONTIRRYLVKGLLIVDGRITQILLALENNHAILVGGFPVSKRVIEHANNQRIPV
MVTHYDTFTVATMINHALSNIRIKTDLKTVEQVNIPTDYGLCEDSSVSEFNTLTKK
TRQVREFVLDYKRVKICGVVSMRDVQDLPTTKLTVMKSNKPIITARPNTSIANISQKMI
FEDLNMLPVTDENNLLGMLTRQAMENLPNHQNNPYTYSQIILSNLETVQYOVV
VEPTMIDSGNMSGVSEFLKEISIRALTQKHKQKIIIEQMMVYFLHAIQIEDELKI
YPLIITENRRSSTIDIEFVDDQVIAKAITTKIN"
complement(5148..5690)
/genes="SPY1356"
complement(5148..5690)
/notes="SPY1356"
/notes="Best Blastp hit = emb|CAA03913.1| (AJ000084)
putative acetyl transferase [Proteus mirabilis]"
/codon_start=1
/transl_table=11
/product="putative acetyl transferase"
/protein_id="AAK34184.1"
/db_xref="GI:13622465"
/translation="MDIWTKLAVFAFFPKVILRPFREDHWDFTSMVNDTKNLVYV
FPEQTKAASDYLLVHSFKFPLGOWAIEDKATHQVIGSIRIEHYDAKTACADIGVFL
NFAFMGGIMTEVYIKLVYLSFHEFGLKTLRIITHLENKASQKVAKKAGPQLKTCFKG
SDRNTKICIKYKNYQLTNDR"
complement(5928..6581)
/genes="grab"
/notes="SPY1357"
complement(5928..6581)
/genes="grab"
/notes="Best Blastp hit = gb|AAD26340.1| (AF124401.1
(AF124401) GRAB precursor [Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="protein GRAB (protein G-related alpha 2W-binding
protein)"
/protein_id="AAK34185.1"
/db_xref="GI:13622466"
/translation="MKKEIKYKCFLRRAAGLVASVLYGTVSAVDSPIQPRII
PNGGTLTNLIGNAPEKLALRNEERAIDELKQAIEDKEATTATEAASSDALEALADOT
DALQSEEAAYVVRADNAADALEALADQTDALQSEEAAYVQSDNAASDAWEKAATPIAL
DVKTKTDTPVVKKEERQVNTLPTTGEESNPFTTAAALAIMVSTGLVYVSSCKEN"
complement(6939..8198)
/genes="murZ"
/notes="SPY1358"
complement(6939..8198)
/genes="murZ"
/notes="Best Blastp hit = sp|P19670|MUA2_BACSU
UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE 2
(ENOYLPRUVATE TRANSFERASE) (UDP-N-ACETYLGLUCOSAMINE
ENOYLPRUVYL TRANSFERASE) (EPT) >gi|2127262|pir|I032354
UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC
2.5.1.7) murZ - Bacillus subtilis
>gi|853767|emb|CAA89875.1| (Z49782)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis] >gi|2636235|emb|CAB15727.1| (Z99122)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis] >gi|2636247|emb|CAB15738.1| (Z99123)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="putative UDP-N-acetylglucosamine
1-carboxyvinyltransferase"
/protein_id="AAK34186.1"

```

Alignment Scores:

ORGANISM Streptococcus pyogenes  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.

REFERENCE 1 (bases 1 to 832)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22):15336-15344 (1999)  
MEDLINE 99269061  
PUBMED 10336419

REFERENCE 2 (bases 1 to 832)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden

FEATURES  
source  
1..832  
/organism="Streptococcus pyogenes"  
/strain="ATCC700294"  
/db\_xref="ATCC:700294"  
/db\_xref="taxon:1314"  
1..99  
sig\_peptide  
CDS  
1..654  
/note="grab"  
/codon\_start=1  
/transl\_table=11  
/product="GRAB precursor"  
/protein\_id="AAD26338.1"  
/db\_xref="GI:4589079"  
/translation="MGKEIKVKFLRRSAAGLVAASVLSVSTVSADVSPTEOPRII  
PNGTLNLLGNAPKALRNEIRAIIDELKQAIIDKEDATTAIAASDADALADQ  
DALQSEAAVVKADNAASDALEALADQDALQSEAEVVDNAASDAWEKAATPIAL  
DVKKTKDPKPVKKBERQNVNLTPTGESNPFTTAALAIWSTGVLSVSSCKEN"  
100..651  
mat\_peptide  
/product="GRAB"  
100..273  
misc\_feature  
/note="Region: alpha2-macroglobulin-binding A domain"  
274..441  
misc\_feature  
/note="Region: repeat motif"  
442..576  
misc\_feature  
/note="Region: cell-wall attachment"  
577..651  
/note="Region: membrane-spanning"  
BASE COUNT 304 a 143 c 169 g 216 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,476-90 Length: 832  
Score: 777.00 Matches: 159  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539a-6 (1-159) x AFI24399 (1-832)

QY 1 ValAspSerProIleGluInProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
|||||  
Db 100 GTTGACTCATTATCAACAGCGCTCGAATATTCCAAATGGCGGACCTTAACATCTT 159  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
|||||  
Db 160 CTTGGCAATGCTCCAGAAAACCTGGCATTACGTAATGAAGAAGAGCCATTGATGAATTA 219  
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
|||||  
Db 220 AAAAACAACGATTGAGGATAAAGAGCTTACGACAGCTATAGAAGCAGCAAGCTTCAGAT 279  
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80  
|||||  
Db 280 GCCTTAGACGATTACGGGATCAACAGAGCGCTTTACATACAGAGAGCGCTGGGTGTT 339  
QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100

Db 340 AAAGCGGATAACGCTGCTAGTGACGCTTAGAAGCATTGGCGGATCAAAACAGACGCTTTA 399  
QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaIleGluLys 120  
|||||  
Db 400 CAATCAGAAGAAGCTGAAGTAGTTCATCAATCAGATACGCTGCTAGTGACGCTGGGAAAA 459  
QY 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140  
|||||  
Db 460 GCAGCACTCCAAATCGCTTTAGATGTTAAGAACTAAAGATACAAAACCTGTAGTTAAA 519  
QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrGlyGluGluSerAsnPro 159  
|||||  
Db 520 AAAGAAGAAGACAAACAGCTTAATACCTCTTACACTGCTGAAGACTCTAAACCCA 576

RESULT 2  
AE006573/c 10029 bp DNA linear BCT 01-JUN-2001  
LOCUS Streptococcus pyogenes M1 GAS strain SF370, section 102 of 167 of  
DEFINITION the complete genome.  
ACCESSION AE006573 AE004092  
VERSION AE006573.1 GI:13622459  
KEYWORDS  
SOURCE Streptococcus pyogenes M1 GAS.  
ORGANISM Streptococcus pyogenes M1 GAS  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 10029)  
AUTHORS Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,  
Primeaux,C., Seate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,  
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,O., Zhu,H., Song,L., White,J.,  
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.  
TITLE Complete genome sequence of an M1 strain of Streptococcus pyogenes  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)  
MEDLINE 21192684  
PUBMED 11296296

REFERENCE 2 (bases 1 to 10029)  
AUTHORS Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,  
Primeaux,C., Seate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,  
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,O., Zhu,H., Song,L., White,J.,  
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.  
TITLE Direct Submission  
JOURNAL Submitted (10-APR-2001) Department of Microbiology and Immunology,  
University of Oklahoma Health Sciences Center, 940 SL Young Blvd,  
Oklahoma City, OK 73104, USA  
FEATURES  
Location/Qualifiers  
1..10029  
/organism="Streptococcus pyogenes M1 GAS"  
/strain="SF370"  
/serotype="M1"  
/db\_xref="taxon:160490"  
complement(135..626)  
/gene="arok"  
/note="SPY1351"  
complement(135..626)  
/gene="arok"  
/EC\_number="2.7.1.71"  
/note="Best Blastp hit = pir|D82693 shikimate kinase  
XF1335 [imported] - Xylella fastidiosa (strain 9a5c)  
>gi19106332|gb|AAAF84144.1|AE003966\_5 (AE003966) shikimate  
kinase [Xylella fastidiosa]"  
/codon\_start=1  
/transl\_table=11  
/product="putative shikimate kinase"  
/protein\_id="AAK34179.1"  
/db\_xref="GI:13622460"  
/translation="MTRKVLGFMGVGKTTVSKHLSMHCKMDMAIEAKIGMSIAAFPE  
QHGIAPFTIESQVLKDLFANDNSIIVTGGVVLGGRNQLLRKHQHNLLVASFE  
TLYORLKHDKKSQRPLFLKYSKEAFYEFYQQRMVFYEGLSGLVIRVDHRTPEVANII  
EGY"  
complement(619..1911)  
/gene="arok"  
/note="SPY1352"

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:12:06 ; Search time 1551.74 Seconds  
(without alignments)  
2144.233 Million cell updates/sec

Title: US-09-847-539A-6  
Perfect score: 777  
Sequence: 1 VDSPIEQPIIPNGGTLTNL.....KKEERQNVNTLTGTGESNP 159

Scoring table: BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DRV=vlh  
-Q/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_092548\_3423/app\_query.fasta\_1.526  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539 -CGN\_1\_1\_1182\_@runat\_10102002\_092548\_3423 -NCFU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FCGAPOP=6 -FCGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl :  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_mu.\*  
19: em\_or.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pi.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgtgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	777	100.0	832	1	AF124399 Streptococcus
2	777	100.0	10029	1	AE006573 Streptococcus
3	760	97.8	468	1	AF124401 Streptococcus
4	754	97.0	469	1	AF124400 Streptococcus
5	536	69.0	804	1	AF124403 Streptococcus
6	534	68.7	717	1	AF124402 Streptococcus
7	155	19.9	1950	1	STRSPGIGP
8	155	19.9	1950	6	IO9115
9	155	19.9	2384	1	SGSPG
10	155	19.9	2384	6	IO8537
11	155	19.9	2456	1	SGPROTG
12	153	19.7	1950	6	IO8536
13	142	18.3	1576	1	SGI48PG
14	126	16.2	969	1	STRABP
15	113	14.5	16727	9	AC020602
16	105	13.5	798	6	IO9107
17	104	13.4	3222	14	HPU23857
18	104	13.4	106516	1	STYPPHMC2
19	99	12.7	48502	7	LAMCG
20	96	12.4	13501	1	AE005258
21	96	12.4	297816	1	AP002553
22	92	11.8	220820	2	AL663060
23	91	11.7	1555	1	STRWAG
24	90	11.6	1469	1	SGI48IGG
25	90	11.6	1469	6	AL2446
26	90	11.6	2022	3	AY060997
27	90	11.6	6484	3	AF335500
28	90	11.6	10432	1	AE005333
29	90	11.6	54784	2	AC017732
30	90	11.6	140973	3	AC008140
31	90	11.6	179669	3	AC008089
32	90	11.6	222605	1	AP002555
33	90	11.6	224795	3	AE003688
34	90	11.6	269726	2	AC020883
35	89	11.5	2437	1	SAG277292
36	89	11.5	9331	1	AE006458
37	89	11.5	20777	2	LMFLCHR34_17
38	89	11.5	110000	2	LMFLCHR34_16
39	89	11.5	291136	1	AP002556
40	88	11.3	848	11	GNS0618W
41	88	11.3	3720	14	S76368
42	88	11.3	43658	14	HSV3PRGEN
43	88	11.3	112930	14	HSGEND
44	88	11.3	189129	2	AC022682
45	88	11.3	256608	10	AL589699

ALIGNMENTS

RESULT 1  
AF124399  
LOCUS  
DEFINITION Streptococcus pyogenes strain ATCC700294 GRAB precursor, gene, complete cds.  
ACCESSION AF124399  
VERSION AF124399.1 GI:4589078  
KEYWORDS Streptococcus pyogenes.  
SOURCE

AF124399 832 bp DNA linear BCT 14-AUG-2000

XX  
SQ Sequence 200 AA;  
Query Match 10.7%; Score 83; DB 16; Length 200;  
Best Local Similarity 54.8%; Pred. No. 0.043;  
Matches 17; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
Qy 128 DVKTKDTKPVVKKFEERQNVNTLPTTGESN 158  
| | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | |  
Db 163 DTKKEDAKKPEAKKEDAKKAETLPTTGESN 193

Search completed: October 13, 2002, 02:08:14  
Job time : 66.6203 secs







PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134766.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140931.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142290.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.



XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 XX N-PSDB; AAS82434.  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 XX Claim 20; SEQ ID No 48606; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 1107 AA:  
 Query Match 11.1%; Score 86; DB 22; Length 1107;  
 Best Local Similarity 26.8%; Pred. No. 0.15;  
 Matches 26; Conservative 16; Mismatches 55; Indels 0; Gaps 0;  
 QY 46 EDKEATTATEAASSDALEALADOTDALQSEEAHVADNAADALEALADOTDALQSEEA 105  
 Db 647 QSSAAASSASAATSETNAKASETSAESKTAASASSASSASSASSASASAKDRTAQSA 706  
 QY 106 EVVQSDNAASDAWEKAATPTALDVKKTKDTPVVKKE 142  
 Db 707 AKGSATTASTKATEAAGSATAAQAQSKSTAESAAATRAE 743  
 RESULT 35  
 AAR71929  
 ID AAR71929 standard; Protein; 664 AA.  
 XX  
 AC AAR71929;  
 XX  
 DT 22-SEP-1995 (first entry)  
 XX  
 DE S. dysgalactiae MIG.  
 XX  
 KW MIG: fast alpha-2-macroglobulin binding protein; FAM;  
 KW plasma proteinase-inhibitor binding protein.  
 XX  
 OS Streptococcus dysgalactiae strain SC1.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..30  
 FT /label= sig\_peptide  
 FT 209..278  
 FT Domain  
 FT /label= IgG1  
 FT /note= "IgG binding domain 1"

FT Domain 279..348  
 FT /label= IgG2  
 FT /note= "IgG binding domain 2"  
 FT 349..418  
 FT /label= IgG3  
 FT /note= "IgG binding domain 3"  
 FT 419..488  
 FT /label= IgG4  
 FT /note= "IgG binding domain 4"  
 FT 489..558  
 FT /label= IgG5  
 FT /note= "IgG binding domain 5"  
 FT 559..638  
 FT /note= "cell wall spanning region"  
 FT 630..635  
 FT /note= "putative wall anchoring motif"  
 FT 639..664  
 FT /note= "membrane spanning domain"  
 XX  
 PN W09507296-A.  
 XX  
 XX 16-MAR-1995.  
 PD  
 XX  
 PF 06-SEP-1994; 94WO-SE00826.  
 XX  
 PR 06-SEP-1993; 93SE-0002855.  
 XX  
 PA (GUSS/) GUSS B.  
 PA (JONS/) JONSSON H.  
 PA (LIND/) LINDBERG M.  
 PA (MUEL/) MUELLER H.  
 PA (RANT/) RANTAMAKI L K.  
 XX  
 PI Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;  
 XX WPI; 1995-123382/16.  
 DR N-PSDB; AAQ89196.  
 XX  
 PT DNA encoding fast alpha 2-macroglobulin-binding proteins - used  
 PT to obtain prods. for sepn., detection or quantification or for  
 PT binding inhibition  
 XX  
 PS Disclosure; Page 31; 50pp; English.  
 XX  
 CC A phage lambda GEM-11 library of S. dysgalactiae SC1 DNA was analyzed  
 CC for fast-acting alpha-2-macroglobulin (FAM)- and IgG-binding  
 CC activity. One clone, which expressed both activities, was  
 CC analyzed to obtain DNA encoding the FAM-binding protein, MIG  
 CC (AAR71929). The mig gene is given in AAQ89196.  
 XX  
 SQ Sequence 664 AA:  
 Query Match 10.9%; Score 85; DB 16; Length 664;  
 Best Local Similarity 56.2%; Pred. No. 0.1;  
 Matches 18; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
 QY 128 DVKTKTDKTPVVKKEERQNVNLTPTTGESNP 159  
 Db 608 DAKKPEAKKEAKKEAKKAATLTPTTGESNP 639  
 RESULT 36  
 ABB58655  
 ID ABB58655 standard; Protein; 489 AA.  
 XX  
 AC ABB58655;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 2757.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

```
XX SQ Sequence 413 AA;
Query Match 11.7%; Score 91; DB 16; Length 413;
Best Local Similarity 59.4%; Pred. No. 0.011;
Matches 19; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

Qy 128 DVKTKDTPVVKKEERONVNTLPTTGESNP 159
   || ||| : ||||| |||
Db 357 DTKKEDAKKPEAKKEAKAATLPTTGESNP 388

RESULT 32
AAP70468
ID AAP70468 standard; Protein; 480 AA.
XX
AC AAP70468;
XX
DT 21-MAY-1991 (first entry)
XX
DE Sequence of polypeptide possessing IgG-binding activity of
DE protein G from Streptococcus G148.
XX
KW Antibody-binding; IgG; IgA; immunoglobulin.
XX
OS Streptococcus G148.
XX
PN W08705631-A.
XX
PD 24-SEP-1987.
XX
PF 20-MAR-1987; 87WO-SE00145.
XX
PR 21-MAR-1986; 86SE-0001325.
XX
PA (PHAA ) PHARMACIA AB.
PA (GUSS/) GUSS B M.
XX
PI Guss BM, Lindberg KM, Flock JI, Uhlen CEM;
XX
DR WPI; 1987-277686/39.
DR N-PSDB; AAN70757.
XX
PT New recombinant DNA molecules - for producing proteins with
PT IgG-binding specificity of protein G or proteins A and G
XX
PS Example; Fig 2: 39pp; English.
XX
CC A recombinant DNA molecule containing a nucleotide sequence which
CC codes for a protein or polypeptide having the same IgG specificity as
CC protein G from Streptococcus G148 (AAN70757) is claimed. See, for
CC example, AAN70754, AAN70755 and AAN70756.
XX
SQ Sequence 480 AA;
Query Match 11.6%; Score 90; DB 8; Length 480;
Best Local Similarity 56.2%; Pred. No. 0.017;
Matches 18; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 128 DVKTKDTPVVKKEERONVNTLPTTGESNP 159
   || ||| : ||||| |||
Db 424 DTKKEDAKKPEAKKEAKAATLPTTGESNP 455

RESULT 33
AAV71049
ID AAV71049 standard; peptide; 19 AA.
XX
AC AAV71049;
XX
DT 29-AUG-2000 (first entry)
XX
DE Streptococcus pyogenes GRAB peptide EKL 18.
```

```
XX GRAB; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection; peptide EKL 18.
XX Streptococcus pyogenes.
OS
PN W0200026240-A2.
PD 11-MAY-2000.
XX
PF 02-NOV-1999; 99WO-GR03631.
XX
PR 02-NOV-1998; 98GE-0023975.
XX
PA (ACTI-) ACTINOVA LTD.
XX
PI Bjorck LH, Rasmussen M;
XX
DR WPI; 2000-365572/31.
XX
XX New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein
PT
XX
PS Example 9; Page 31; 67pp; English.
XX
CC The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a GRAB peptide EKL 18 useful in vaccine composition. It was used
CC to produce immunoglobulin G (IgG) antibodies specific for native GRAB
CC protein in sheep. The peptide has a cysteine insert at the C-terminal
CC for attachment to a hetero-bifunctional linker.
XX
SQ Sequence 19 AA;
Query Match 11.1%; Score 86; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 EKALRNEERAIDELKKQ 43
   ||||| ||||| |||||
Db 1 EKALRNEERAIDELKKQ 18

RESULT 34
ABG18247
ID ABG18247 standard; Protein; 1107 AA.
XX
AC ABG18247;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #18238.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN W0200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
```





PR 19-JUN-1987; 87US-0063959.  
PR 20-JUN-1988; 88US-0209236.  
PR 19-JUN-1990; 90US-0540169.  
PR 21-APR-1992; 92US-0871539.  
XX  
PA (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.  
XX  
XX Fahnstock SR;  
XX  
XX WPI: 1994-159179/19.  
DR N-PSDB; AAQ64648.  
XX  
XX New recombinant streptococcal protein G variants - useful for  
PT antibody detection and purification and for therapy  
XX  
XX Claim 3: Column 46; 48pp; English.  
PS  
XX Protein G isolated from Streptococcus has IgG-binding activity which  
CC has been localised to the B repeating structure (see AAR53290).  
CC The sequence AAR53294 represents a claimed Streptococcal Protein G  
CC variant comprising the B domain active site and retaining  
CC IgG-binding activity.  
XX  
XX Sequence 265 AA;  
Query Match 13.5%; Score 105; DB 15; Length 265;  
Best Local Similarity 69.7%; Pred. No. 0.00012;  
Matches 23; Conservative 1; Mismatches 9; Indels 0; Gaps 0;  
QY 4 PIERPRIINGGTLTLNGLNAPEKALRNEERA 36  
||| || ||| |||||: ||||| |  
Db 34 PIEDTIIRNGGELTNLGNSETTLALRNEESA 66  
RESULT 26  
AAY71051  
ID AAY71051 standard; peptide; 20 AA.  
XX  
AC AAY71051;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes GRAB peptide KKT 19.  
XX  
KW GRAB; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection; peptide KKT 19.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI: 2000-365572/31.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
to group A streptococcus and purifying the binding protein -  
XX  
PS Example 9; Page 31; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived

CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a GRAB peptide KKT 19 useful in vaccine composition. It was used  
CC to produce immunoglobulin G (IgG) antibodies specific for native GRAB  
CC protein in sheep. The peptide has a cysteine insert at the C-terminal  
CC for attachment to a hetero-bifunctional linker.  
XX  
XX Sequence 20 AA;  
Query Match 12.5%; Score 97; DB 21; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5e-05;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 130 KTKDTPVVKKEERQNVN 148  
||||| ||||| ||||| |||||  
Db 1 KTKDTPVVKKEERQNVN 19  
RESULT 27  
AAY71047  
ID AAY71047 standard; peptide; 19 AA.  
XX  
AC AAY71047;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes GRAB peptide DSP 18.  
XX  
KW GRAB; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection; peptide DSP 18.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI: 2000-365572/31.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
to group A streptococcus and purifying the binding protein -  
XX  
PS Example 9; Page 31; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a GRAB peptide DSP 18 useful in vaccine composition. It was used  
CC to produce immunoglobulin G (IgG) antibodies specific for native GRAB  
CC protein in sheep. The peptide has a cysteine insert at the C-terminal  
CC for attachment to a hetero-bifunctional linker.  
XX  
XX Sequence 19 AA;  
Query Match 12.4%; Score 96; DB 21; Length 19;  
Best Local Similarity 100.0%; Pred. No. 6.3e-05;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```
Matches 23; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 4 PIEQPIIPNGGTLTLNGLNAPEKALRNEERA 36
   III III IIIIIII: IIIIIII I
Db 4 PIEDPIIRNGGELTNLGNSETTLALRNEESA 36

RESULT 23
AAR07004
ID AAR07004 standard; protein; 235 AA.
XX
AC AAR07004;
XX
DT 17-JAN-1991 (first entry)
XX
DE Protein G variant.
XX
KW Immunoglobulin.
XX
OS Streptococcus sp. Lancefield Group G strain.
XX
PN US4956296-A.
XX
PD 11-SEP-1990.
XX
PF 20-JUN-1988; 88US-0209236.
XX
PR 20-JUN-1988; 88US-0209236.
XX
PR 14-FEB-1986; 86US-0829354.
XX
PR 23-APR-1986; 86US-0854887.
XX
PR 17-FEB-1987; 87WO-US00329.
XX
PR 19-JUN-1987; 87US-0063959.
XX
PA (GENE-) GENEX CORP.
XX
PI Farnestock SR;
XX
WPI; 1990-297491/39.
XX
N-PSDB; AAQ06009.
XX
Recombinant Protein G variants - obtd. using a cloned gene
PT encoding Protein G from Streptococcus sp., used for binding
PT immunoglobulin.
XX
PS Disclosure; Column 9-16; 48pp; English.
XX
Sequence may be incorporated into a non-pathogenic host eg. E.coli,
CC where they may be expressed at high levels. The proteins have a
CC higher binding efficiency and capacity for immunoglobulin, and may
CC be used for purifying, detecting and isolating antibodies.
XX
SQ Sequence 235 AA;
Query Match 13.5%; Score 105; DB 11; Length 235;
Best Local Similarity 69.7%; Pred. No. 0.0001;
Matches 23; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 4 PIEQPIIPNGGTLTLNGLNAPEKALRNEERA 36
   III III IIIIIII: IIIIIII I
Db 4 PIEDPIIRNGGELTNLGNSETTLALRNEESA 36

RESULT 24
AAR10011
ID AAR10011 standard; protein; 265 AA.
XX
AC AAR10011;
XX
DT 13-MAR-1991 (first entry)
XX
DE Type 4 Gx7809 protein G variant.
XX
KW Immunoglobulins; Ig.
```

```
XX
OS Streptococcus sp Gx7809.
XX
PW US4977247-A.
XX
PD 11-DEC-1990.
XX
PF 19-MAY-1989; 89US-0354264.
XX
PR 19-MAY-1989; 89US-0354264.
XX
PR 14-FEB-1986; 86US-0829354.
XX
PR 23-APR-1986; 86US-0854887.
XX
PR 17-FEB-1987; 87WO-US00329.
XX
PR 19-JUN-1987; 87US-0063959.
XX
PR 20-JUN-1988; 88US-0209236.
XX
PA (GENE-) GENEX CORP.
XX
PI Farnestock SR, Lee T, Wroble MH;
XX
WPI; 1991-006758/01.
XX
Q-PSDB; Q10007.
XX
Immoblilised protein G variants - used for detection, isolation
PT and purificn. immunoglobulin(s) and immunoglobulin fragments
PT
XX
PS Disclosure; Column 11; 52pp; English.
XX
Protein G variant product carries active regions B1 and B2, it may
CC be immobilised and exhibits different binding profiles. The bound
CC protein is useful in purification and detection of Igs and fragments.
XX
SQ Sequence 265 AA;
Query Match 13.5%; Score 105; DB 12; Length 265;
Best Local Similarity 69.7%; Pred. No. 0.00012;
Matches 23; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 4 PIEQPIIPNGGTLTLNGLNAPEKALRNEERA 36
   III III IIIIIII: IIIIIII I
Db 34 PIEDPIIRNGGELTNLGNSETTLALRNEESA 66

RESULT 25
AAR53294
ID AAR53294 standard; Protein; 265 AA.
XX
AC AAR53294;
XX
DT 06-JAN-1995 (first entry)
XX
DE IgG-binding Streptococcus Protein G variant.
XX
KW Streptococcus Protein G; variant; IgG binding activity;
KW immunoglobulin; Lancefield Group G; bacterial Fc receptor.
XX
OS Streptococcus sp.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT Protein /label= secretion_sequence
FT /label= 31..265
FT /label= Protein_G_variant
FT /note= "claimed without the secretion sequence"
XX
PN US5312901-A.
XX
XX
PD 17-MAY-1994.
XX
PF 14-FEB-1986; 86US-0829354.
XX
PR 14-FEB-1986; 86US-0829354.
XX
PR 23-APR-1986; 86US-0854887.
```

```
XX WC200026240-A2.
PN
PD 11-MAY-2000.
XX
PF 02-NOV-1999; 99WO-G803631.
XX
PR 02-NOV-1998; 98GB-0023975.
XX
PA (ACTI-) ACTINOVA LTD.
XX
PI Bjorck LH, Rasmussen M;
XX
DR WPI; 2000-365572/31.
XX
PT New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein
XX
PS Claim 1; Page 55; 67pp; English.
XX
CC The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding
CC to residues 34-56. This fragment is capable of binding alpha2M
CC and useful in vaccine composition.
XX
SQ Sequence 23 AA;
Query Match 15.4%; Score 120; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 8.7e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VDSPIEQPRIIPNGGTLNLLGN 23
DB 1 VDSPIEQPRIIPNGGTLNLLGN 23
RESULT 21
AAY71048
ID AAY71048 standard; peptide; 25 AA.
XX
XX AAY71048;
XX
XX 29-AUG-2000 (first entry)
XX
XX Streptococcus pyogenes GRAB peptide EKL 24.
DE
XX GRAB; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antitoxic;
KW immune response; Streptococcus pyogenes infection; peptide EKL 24.
XX
XX Streptococcus pyogenes.
OS
XX WC200026240-A2.
PN
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-G803631.
PF
XX
PR 02-NOV-1998; 98GB-0023975.
XX
PA (ACTI-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
PI
XX WPI; 2000-365572/31.
DR
XX
```

```
PT New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein
XX
XX Example 9; Page 31; 67pp; English.
XX
CC The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a GRAB peptide EKL 24 useful in vaccine composition. It was used
CC to produce immunoglobulin G (IgG) antibodies specific for native GRAB
CC protein in sheep. The peptide has a cysteine insert at the C-terminal
CC for attachment to a hetero-bifunctional linker.
XX
SQ Sequence 25 AA;
Query Match 14.8%; Score 115; DB 21; Length 25;
Best Local Similarity 100.0%; Pred. No. 4e-07;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 EKALRNEERAIDELKKQAIEDKE 49
DB 1 EKALRNEERAIDELKKQAIEDKE 24
RESULT 22
AAP94785
ID AAP94785 standard; protein; 235 AA.
XX
XX AAP94785;
XX
XX 04-JUL-1990 (first entry)
DT
XX Protein G variant.
DE
XX Protein G; immunoglobulin; Fc receptor; ds.
KW
XX Streptococcus sp.
OS
XX WO8810306-A.
PN
XX 29-DEC-1988.
XX
XX 20-JUN-1988; 88WO-US02084.
PF
XX 19-JUN-1987; 87US-0063959.
PR
XX (GENE-) GENEX CORP.
XX
XX Fahnstock SR;
XX
XX WPI; 1989-023848/03.
DR
XX N-PSDB; AAN91099.
XX
XX Cloned protein G variant genes -
XX expressing proteins having immunoglobulin-binding properties of
XX protein G and derived from Streptococcus sp.
XX
XX Claim 17; Page 88; 116pp; English.
PS
XX
XX Gene for protein G variant of non-pathogenic streptococcus sp. allowing
CC isolation of the protein and variants, useful as bacterial Fc receptors
CC eg in purification and detection of Abs., screening of hybridoma clones
CC and treatment of disease.
XX
XX
SQ Sequence 235 AA;
Query Match 13.5%; Score 105; DB 10; Length 235;
Best Local Similarity 69.7%; Pred. No. 0.0001;
```

XX AAV57611;  
 AC 10-MAR-2000 (first entry)  
 DT Streptococcus strain G 148 protein.  
 DE  
 XX Streptococcus strain G 148; protein G'; protein G primer; liposome;  
 KW liposomal delivery complex; connecting moiety; antibody; FC region;  
 KW diagnosis; therapeutic agent; antibiotic; antipressant; antiviral;  
 KW antitumorigenic; cytokine; hormone; imaging agent; neurotransmitter;  
 KW stimulant; cytotoxic agent; malignant cell.  
 XX OS Streptococcus sp.  
 XX WO9959643-A2.  
 PN 25-NOV-1999.  
 XX PD  
 XX 19-MAY-1999; 99WO-US11177.  
 XX PF  
 XX 20-MAY-1998; 98US-0086347.  
 XX PR  
 XX (SDGS-) SDG INC.  
 XX PA  
 XX Lau JR;  
 XX WPI; 2000-062383/05.  
 DR N-PSDB; AAZ47931.  
 XX New liposome constructs comprising a liposome connected to an antibody,  
 PT used, e.g. for delivery of cytotoxic agents to malignant cells -  
 XX  
 PS Disclosure: Fig 1; 32pp; English.  
 XX  
 CC The present invention describes a construct for connecting an antibody  
 CC or antibody fragment to a liposome. The construct comprises protein G'  
 CC and a linking moiety for connecting the protein G' to the liposome.  
 CC Also described are: (1) a liposomal delivery vehicle comprising: (a) a  
 CC liposome; and (b) a connecting moiety connected to the liposome, which  
 CC specifically binds the FC region of an antibody, for connecting the  
 CC antibody to the liposome; and (2) forming a liposomal carrier comprising:  
 CC (a) providing liposomal components having binding moieties, for forming  
 CC a liposome; (b) providing a construct comprising a linking moiety and a  
 CC connecting moiety bound together; and (c) combining the liposomal  
 CC components and the construct and sonicating the combination so that the  
 CC binding moieties are exposed on a surface of the liposome, for binding to  
 CC the linking moiety, and to facilitate the binding between them. The  
 CC products can be used for the delivery of diagnostic or therapeutic  
 CC agents. The liposomes may contain or may be associated with a diagnostic  
 CC or therapeutic agent, e.g. antibiotics, antipressants,  
 CC antitumorigenics, antivirals, cytokines, hormones, imaging agents,  
 CC neurotransmitters, or stimulants. They can be used particularly for the  
 CC delivery of cytotoxic agents to malignant cells. The protein G'  
 CC connecting moiety provides a liposomal delivery complex having improved  
 CC targeting efficiency. As a result of the binding between protein G' and  
 CC the FC region of antibodies, protein G' shields the FC regions of the  
 CC attached antibodies from non-specific binding to cell-surfaces, other  
 CC proteins, and anatomical structures. The present sequence represents  
 CC an unidentified protein encoded by the same sequence which encodes  
 CC protein G'.  
 XX  
 SQ Sequence 103 AA;  
 Query Match 18.3%; Score 142; DB 21; Length 103;  
 Best Local Similarity 57.9%; Pred. No. 1e-09;  
 Matches 33; Conservative 3; Mismatches 21; Indels 0; Gaps 0;  
 QY 1 VDSPIEQPIIPNGGTLTNLGNAPKALNRNEERAIKDKQAIEDKATTAIEAA 57  
 DB 34 VDSPIEDTPIIRNGGELTNLGNSETLALNRNEESATADLTAAAVADTVAAAAAENA 90

RESULT 19  
 AAY71038  
 ID AAY71038 standard; peptide; 28 AA.  
 XX  
 AC AAY71038;  
 DT 29-AUG-2000 (first entry)  
 XX Streptococcus pyogenes strain SF370 GRAB protein fragment #3.  
 DE  
 XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
 KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
 KW immune response; Streptococcus pyogenes infection.  
 XX OS Streptococcus pyogenes.  
 XX WO200026240-A2.  
 PN 11-MAY-2000.  
 XX PD  
 XX 02-NOV-1999; 99WO-GB03631.  
 XX PF  
 XX 02-NOV-1998; 98GB-0023975.  
 XX PR  
 XX (ACTI-) ACTINOVA LTD.  
 XX PA  
 XX Bjorck LH, Rasmussen M;  
 XX WPI; 2000-365572/31.  
 XX DR  
 XX New alpha2M binding protein for generating a protective immune response  
 PT to group A streptococcus and purifying the binding protein -  
 XX  
 PS Claim 3; Page 56; 67pp; English.  
 XX  
 CC The patent discloses a new family of proteins termed GRAB (protein G  
 CC related alpha2M binding protein) from Streptococcus pyogenes which have  
 CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
 CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2M from a sample. The present sequence  
 CC is one of the repeat regions of GRAB protein from  
 CC S. pyogenes strain SF370 corresponding to residues 92-119. This fragment  
 CC is useful in vaccine composition.  
 XX  
 SQ Sequence 28 AA;  
 Query Match 16.6%; Score 129; DB 21; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-09;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 SDALALADQTDALQSEEAAYVVKADNAA 86  
 DB 1 SDALALADQTDALQSEEAAYVVKADNAA 28  
 RESULT 20  
 AAY71036  
 ID AAY71036 standard; peptide; 23 AA.  
 XX  
 AC AAY71036;  
 DT 29-AUG-2000 (first entry)  
 XX Streptococcus pyogenes strain SF370 GRAB protein fragment #1.  
 DE  
 XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
 KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
 KW immune response; Streptococcus pyogenes infection.  
 XX OS Streptococcus pyogenes.

FT Region /note= "involved in IgG binding activity"  
 358..372  
 /label= b  
 FT Active-site /note= "linking region"  
 373..427  
 /label= B3  
 FT Misc-difference /note= "involved in IgG binding activity"  
 376  
 /note= "corresponds to CCT codon"  
 428..442  
 /label= b  
 FT Region /note= "linking region"  
 443..497  
 /label= B2  
 FT Active-site /note= "involved in IgG binding activity"  
 466  
 /note= "corresponds to ACT codon"  
 531..535  
 /label= C1  
 FT Region /note= "linking region"  
 536..540  
 /label= C2  
 FT Region /note= "linking region"  
 541..545  
 /label= C3  
 FT Region /note= "linking region"  
 546..550  
 /label= C4  
 FT Region /note= "linking region"  
 551..555  
 /label= C5  
 FT Misc-difference 592  
 /note= "corresponds to GAA codon"

US5312901-A.  
 17-MAY-1994.  
 14-FEB-1986; 86US-0829354.  
 14-FEB-1986; 86US-0829354.  
 23-APR-1986; 86US-0854887.  
 19-JUN-1987; 87US-0063959.  
 20-JUN-1988; 88US-0209236.  
 19-JUN-1990; 90US-0540169.  
 21-APR-1992; 92US-0871539.  
 (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.  
 Fahnestock SR;  
 WPI; 1994-159179/19.  
 N-PSDB; AAQ75036.

PT New recombinant streptococcal protein G variants - useful for  
 antibody detection and purification and for therapy

Example 5; Fig 9; 48pp; English.

CC A 2.4kb HindIII fragment containing the entire coding sequence for  
 CC Protein G was isolated from Streptococcus GX7805 using the 1.9kb  
 CC Protein G coding sequence from Streptococcus GX7809. The Protein G  
 CC has IgG-binding activity which has been localised to the B repeating  
 CC structure. Streptococcal Protein G variants comprising the B domains  
 CC are claimed.

XX Sequence 593 AA;  
 Query Match 19.9%; Score 155; DB 15; Length 593;  
 Best Local Similarity 56.1%; Pred. No. 2.2e-10;  
 Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 VDSPTEQPIRPNNGTTLNLGNAPKALRNEERAIDELKKQAIEDKENTTAIEAASSD 60  
 DB 34 VDSPIEDTPIIRNGGELTTLNLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93  
 QY 61 ALEALA 66

Db 94 AWEAAA 99  
 RESULT 17  
 AAR10005  
 ID AAR10005 standard; protein; 594 AA.  
 XX AAR10005;  
 AC AAR10005;  
 DT 13-MAR-1991 (first entry)  
 XX Streptococcus GX7805 protein G.  
 DE Streptococcus GX7805 protein G.  
 XX Streptococcus GX7805 protein G.  
 KW Immunoglobulins; Ig.  
 XX Streptococcus sp GX7805.  
 OS Streptococcus sp GX7805.  
 FH Key Location/Qualifiers  
 FT Active-site 304..358  
 /label= Active Site B1  
 FT Active-site 374..428  
 /label= Active Site B3  
 FT Active-site 444..498  
 /label= Active Site B2  
 XX US4977247-A.  
 PN 11-DEC-1990.  
 PD 19-MAY-1989; 89US-0354264.  
 PF 19-MAY-1989; 89US-0354264.  
 PR 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 17-FEB-1987; 87WO-US00329.  
 PR 19-JUN-1987; 87US-0063959.  
 PR 20-JUN-1988; 88US-0209236.  
 XX (GENE-) GENEX CORP.  
 XX Fahnestock SR, Lee T, Wroble MH;  
 WPI; 1991-006758/01.  
 DR Q-PSDB; Q10002.  
 XX Immobilised protein G variants - used for detection, isolation  
 PT and purificn. immunoglobulin(s) and immunoglobulin fragments  
 XX Disclosure; Fig 9; 52pp; English.  
 PS Protein G gene product may be modified allowing the variant to  
 CC be immobilised and exhibit different binding profiles. The bound  
 CC protein is useful in purification and detection of Igs and fragments.  
 CC  
 XX Sequence 594 AA;  
 Query Match 19.9%; Score 155; DB 12; Length 594;  
 Best Local Similarity 56.1%; Pred. No. 2.2e-10;  
 Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 VDSPTEQPIRPNNGTTLNLGNAPKALRNEERAIDELKKQAIEDKENTTAIEAASSD 60  
 DB 34 VDSPIEDTPIIRNGGELTTLNLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93  
 QY 61 ALEALA 66  
 DB 94 AWEAAA 99  
 RESULT 18  
 AAY57611  
 ID AAY57611 standard; Protein; 103 AA.

```

XX PD 17-MAY-1994.
XX PF 14-FEB-1986; 86US-08293354.
XX PR 14-FEB-1986; 86US-08293354.
XX PR 23-APR-1986; 86US-0854887.
XX PR 19-JUN-1987; 87US-0063959.
XX PR 20-JUN-1988; 88US-0209236.
XX PR 19-JUN-1990; 90US-0540169.
XX PR 21-APR-1992; 92US-0871539.
XX PA (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX PI Fahnstock SR;
XX DR WPI; 1994-159179/19.
XX DR N-PSDB; AAQ64644.
XX PS New recombinant streptococcal protein G variants - useful for
XX antibody detection and purification and for therapy
XX PS Example 2: Fig 3 and Fig 8; 48pp; English.
XX CC A 1.9kb HindIII fragment containing the entire coding sequence for
XX protein G was isolated from Streptococcus GX7809. The Protein G has
XX IgG-binding activity which has been localised to the B repeating
XX structure. Streptococcal Protein G variants comprising the B domains
XX are claimed.
XX SQ Sequence 448 AA;
Query Match 19.9%; Score 155; DB 15; Length 448;
Best Local Similarity 56.1%; Pred. No. 1.5e-10;
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;
QY 1 VDSPIEQPRIIPNGTGLTNLIGNAPEKIALRNEERAIDELKKQAIEDKEATTAIEAASD 60
DB 34 VDSPIEDPIIRNGGELTNLIGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93
QY 61 ALEALA 66
DB 94 AWEAAA 99
RESULT 15
AAR07014
XX ID AAR07014 standard; protein; 593 AA.
XX AC AAR07014;
XX DT 17-JAN-1991 (first entry)
XX DE Streptococcus sp. Lancefield Group G strain.
XX DE Protein G variant with three active sites.
XX KW Immunoglobulin.
XX OS Streptococcus sp. Lancefield Group G strain.
XX FH Key Location/Qualifiers
XX FT Active-site 303..372
XX FT Active-site /label=B1
XX FT Active-site 373..427
XX FT Active-site /label=B3
XX FT Active-site 443..497
XX FT Active-site /label=B2
XX FT US4956296-A.
XX PN 11-SEP-1990.
XX PD 20-JUN-1988; 88US-0209236.
XX PF
XX

```

```

PR 20-JUN-1988; 88US-0209236.
PR 14-FEB-1986; 86US-08293354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87WO-US00329.
PR 19-JUN-1987; 87US-0063959.
XX (GENE-) GENEX CORP.
XX PA Fahnstock SR;
XX PI WPI; 1990-297491/39.
XX DR N-PSDB; AAQ06019.
XX PS Recombinant Protein G variants - obtd. using a cloned gene
XX encoding Protein G from Streptococcus sp., used for binding
XX immunoglobulin.
XX PS Disclosure; Fig 9; 48pp; English.
XX CC Fragments and variants of the sequence are claimed esp. where
XX incorporated into a non-pathogenic host eg. E.coli, and expressed
XX at high levels.
XX CC The variants have a higher binding efficiency and capacity for
XX immunoglobulin, and may be used for purifying, detecting and
XX isolating antibodies.
XX SQ Sequence 593 AA;
Query Match 19.9%; Score 155; DB 11; Length 593;
Best Local Similarity 56.1%; Pred. No. 2.2e-10;
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;
QY 1 VDSPIEQPRIIPNGTGLTNLIGNAPEKIALRNEERAIDELKKQAIEDKEATTAIEAASD 60
DB 34 VDSPIEDPIIRNGGELTNLIGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93
QY 61 ALEALA 66
DB 94 AWEAAA 99
RESULT 16
AAR62944
XX ID AAR62944 standard; Protein; 593 AA.
XX AC AAR62944;
XX DT 10-JAN-1995 (first entry)
XX DE Streptococcus Protein G derived from strain GX7805.
XX DE Streptococcus Protein G; variant; IgG binding activity;
XX KW immunoglobulin; Lancefield Group G; bacterial Fc receptor.
XX OS Streptococcus sp. GX7805.
XX FH Key Location/Qualifiers
XX FT Domain 106..140
XX FT Region /label=A1
XX FT Region 141..178
XX FT Domain /label=a1
XX FT Domain /note="linking region"
XX FT Region 179..215
XX FT Region /label=A2
XX FT Region 216..253
XX FT Domain /label=a2
XX FT Domain /note="linking region"
XX FT Misc-difference 269
XX FT Active-site /label=A3
XX FT Active-site /note="corresponds to GGA codon"
XX FT Active-site 303..357
XX FT Active-site /label=B1

```

KW Immunoglobulin.  
 XX Streptococcus sp. Lancefield Group G strain.  
 XX  
 FH Key Location/Qualifiers  
 FT Active-site 228..282  
 FT /label=B1  
 FT Active-site 298..352  
 FT /label=B2  
 XX  
 PN US4956296-A.  
 XX  
 XX 11-SEP-1990.  
 PD  
 XX 20-JUN-1988; 88US-0209236.  
 PF  
 XX 20-JUN-1988; 88US-0209236.  
 PR 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 17-FEB-1987; 87WO-US00329.  
 PR 19-JUN-1987; 87US-0063959.  
 XX  
 PA (GENE-) GENEX CORP.  
 XX  
 XX Fahnestock SR;  
 XX WPI; 1990-297491/39.  
 DR N-PSDB; AAQ06018.  
 XX  
 PT Recombinant Protein G variants - obtd. using a cloned gene  
 PT encoding Protein G from Streptococcus sp., used for binding  
 PT immunoglobulin.  
 PS Disclosure; Fig 8a-c; 48pp; English.  
 XX  
 CC. Fragments and variants of the sequence are claimed esp. where  
 CC incorporated into a non-pathogenic host eg. E.coli, and expressed  
 CC at high levels.  
 CC The variants have a higher binding efficiency and capacity for  
 CC immunoglobulin, and may be used for purifying, detecting and  
 CC isolating antibodies.  
 XX.  
 SQ Sequence 448 AA;  
 Query Match 19.9%; Score 155; DB 11; Length 448;  
 Best Local Similarity 56.1%; Pred. No. 1.5e-10;  
 Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 VDSPIEQPRIIPNGGTLTLLGNAPKALRNEERAIDELKKQAIEDKEATTAEASDD 60  
 DB 34 VDSPIEDTPIIRNGGELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93  
 QY 61 ALEALA 66  
 DB 94 AWEAAA 99  
 DE  
 RESULT 13  
 AARI0004  
 ID AARI0004 standard; Protein; 448 AA.  
 XX  
 AC AARI0004;  
 XX  
 DT 13-MAR-1991 (first entry)  
 XX  
 DE Streptococcus GX7809 protein G.  
 XX  
 KW Immunoglobulins; Ig.  
 XX  
 OS Streptococcus sp GX7809.  
 XX  
 FH Key Location/Qualifiers  
 FT Active-site 228..282

FT Active-site /label= Active site B1  
 FT 298..352  
 FT /label= Active site B2  
 XX  
 PN US4977247-A.  
 XX  
 XX 11-DEC-1990.  
 PD  
 XX 19-MAY-1989; 89US-0354264.  
 PF  
 XX 19-MAY-1989; 89US-0354264.  
 PR 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 17-FEB-1987; 87WO-US00329.  
 PR 19-JUN-1987; 87US-0063959.  
 PR 20-JUN-1988; 88US-0209236.  
 XX  
 PA (GENE-) GENEX CORP.  
 XX  
 XX Fahnestock SR, Lee T, Wroble MH;  
 XX WPI; 1991-006758/01.  
 DR Q-PSDB; Q10001.  
 XX  
 PT Immobilised protein G variants - used for detection, isolation  
 PT and purificn. immunoglobulin(s) and immunoglobulin fragments  
 XX  
 PS Disclosure; Fig 8; 52pp; English.  
 XX  
 CC Protein G gene product may be modified allowing the variant to be  
 CC immobilised and exhibit different binding profiles. The bound  
 CC protein is useful in purification and detection of Igs and fragments.  
 XX  
 SQ Sequence 448 AA;  
 Query Match 19.9%; Score 155; DB 12; Length 448;  
 Best Local Similarity 56.1%; Pred. No. 1.5e-10;  
 Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 VDSPIEQPRIIPNGGTLTLLGNAPKALRNEERAIDELKKQAIEDKEATTAEASDD 60  
 DB 34 VDSPIEDTPIIRNGGELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93  
 QY 61 ALEALA 66  
 DB 94 AWEAAA 99  
 DE  
 RESULT 14  
 AAR53290  
 ID AAR53290 standard; Protein; 448 AA.  
 XX  
 AC AAR53290;  
 XX  
 DT 06-JAN-1995 (first entry)  
 XX  
 DE Streptococcus Protein G derived from strain GX7809.  
 XX  
 KW Streptococcus Protein G; variant; IgG binding activity;  
 KW immunoglobulin; Lancefield Group G; bacterial Fc receptor.  
 XX  
 OS Streptococcus sp. GX7809.  
 XX  
 FH Key Location/Qualifiers  
 FT Active-site 228..282  
 FT /label= B1  
 FT Region 283..297  
 FT /label= b  
 FT /note="linking region"  
 FT Active-site 298..352  
 FT /label= B2  
 XX  
 PN US5312901-A.

CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
 CC to residues 34-91. This fragment is capable of binding alpha2M  
 CC and useful in vaccine composition.

XX Sequence 58 AA;

Query Match 36.7%; Score 285; DB 21; Length 58;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-27;  
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDSPIEQPRIIPNGGTLNLLGNAPKALRNEERATDELKKAIEDKEATTAEAS 58  
 Db 1 VDSPIEQPRIIPNGGTLNLLGNAPKALRNEERATDELKKAIEDKEATTAEAS 58

# RESULT 10

AAP70493  
 ID AAP70493 standard; protein; 448 AA.

XX AAP70493;

XX 06-MAR-1991 (first entry)

XX Protein G.

XX Protein G; antibody; Fc receptor;

XX Streptococcus Lancefield Group G strain.

XX Key Location/Qualifiers

FT Active-site 228..297

FT Active-site /label=active site B1

FT Active-site 298..352

FT Active-site /label=active site B2

XX W08705025-A.

XX 27-AUG-1987.

XX 17-FEB-1987; 87WO-US00329.

XX 22-APR-1986; 86US-0854997.

XX 14-FEB-1986; 86US-0829334.

XX (GENE-) GENEX CORP.

XX (FAHN/) FAHNESTOCK S R.

XX Fahnestock S;

XX WPI; 1987-250197/35.

XX N-PSDB; AAN70811.

XX Cloned Protein G gene - used for producing Protein G for  
 XX detection and purification of antibodies and treatment of diseases

XX Disclosure; Fig. 3; 68pp; English.

XX Protein G expressed by inserting the gene into an expression  
 XX vector. A second vector may also be used as a cryptic helper plasmid  
 XX to stably maintain the first plasmid in the host cell. Bacterial Fc  
 XX receptors such as protein G can be used to detect and purify  
 XX antibodies, and in the treatment of disease. Fc receptors are useful  
 XX to purify antibodies to be used in the purification of protein drugs and  
 XX as therapeutics. High levels of protein G can be obtained in conditions  
 XX favourable for isolation, using a non-pathogenic host. Suitable  
 XX cloning vectors are lambda gt11, M13mp9 and pC1066.

XX Sequence 448 AA;

Query Match 19.9%; Score 155; DB 8; Length 448;

Best Local Similarity 56.1%; Pred. No. 1.5e-10;

Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 VDSPIEQPRIIPNGGTLNLLGNAPKALRNEERATDELKKAIEDKEATTAEAS 60  
 Db 34 VDSPIEQPRIIPNGGTLNLLGNAPKALRNEERATDELKKAIEDKEATTAEAS 93

QY 61 ALEALA 66

Db 94 AWEAAA 99

# RESULT 11

AAP95030

ID AAP95030 standard; protein; 448 AA.

XX AAP95030;

XX 04-JUL-1990 (first entry)

XX Protein G.

XX Protein G; immunoglobulin; Fc receptor; ds.

XX Streptococcus sp.

XX W08810306-A.

XX 29-DEC-1988.

XX 20-JUN-1988; 88WO-US02084.

XX 19-JUN-1987; 87US-0063959.

XX (GENE-) GENEX CORP.

XX Fahnestock SR;

XX WPI; 1989-023848/03.

XX N-PSDB; AAN91093.

XX Cloned protein G variant genes -  
 XX expressing proteins having immunoglobulin-binding properties of  
 XX protein G and derived from Streptococcus sp.

XX Disclosure; 116pp; English.

XX Protein G of non-pathogenic streptococcus and variants may be isolated,  
 XX useful as bacterial Fc receptors eg in purification and detection of Abs.  
 XX screening of hybridoma clones and treatment of disease.

XX Sequence 448 AA;

Query Match 19.9%; Score 155; DB 10; Length 448;

Best Local Similarity 56.1%; Pred. No. 1.5e-10;

Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 VDSPIEQPRIIPNGGTLNLLGNAPKALRNEERATDELKKAIEDKEATTAEAS 60

Db 34 VDSPIEQPRIIPNGGTLNLLGNAPKALRNEERATDELKKAIEDKEATTAEAS 93

# RESULT 12

AAR07013

ID AAR07013 standard; protein; 448 AA.

XX AAR07013;

XX 17-JAN-1991 (first entry)

XX Protein G variant with two active sites.

PT New alpha2M binding protein for generating a protective immune response  
to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 5; Page 61-62; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
related alpha2M binding protein) from Streptococcus pyogenes which have  
the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
is a partial GRAB protein from S. pyogenes strain AP49.  
CC The protein has alpha2M binding region and is useful in vaccine  
composition.  
XX  
SQ Sequence 271 AA;  
Query Match 69.0%; Score 536; DB 21; Length 271;  
Best Local Similarity 92.6%; Pred. No. 9.4e-58;  
Matches 113; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
QY 1 VDSPIEQPRIIPNGGTLTLLGNAPKALRNEERAIDELKKQAIEDKEATTAIEAASSD 60  
DB 4 VDSPIEQPRIIPNGGTLTLLGNAPKALRNEERAIDELKKQAIEDKEATTAIEAASSD 63  
QY 61 ALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVVKADNAASDAWEK 120  
DB 64 ALEALADQADALQSEEAAYVQSDNAASDALEALADQADALQSEEAAYVQSDNAAGDALEA 123  
QY 121 AA 122  
DB 124 LA 125  
RESULT 8  
AAV71043  
ID AAV71043 standard; Protein; 259 AA.  
XX  
AC AAV71043;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain KTL9 partial GRAB protein.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
DR N-PSDB; AAD00561.  
XX  
CC New alpha2M binding protein for generating a protective immune response  
to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 5; Page 59-60; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G

CC related alpha2M binding protein) from Streptococcus pyogenes which have  
the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
is a partial GRAB protein from S. pyogenes strain KTL9.  
CC The protein has alpha2M binding region and is useful in vaccine  
composition.  
XX  
SQ Sequence 259 AA;  
Query Match 68.7%; Score 534; DB 21; Length 259;  
Best Local Similarity 91.8%; Pred. No. 1.6e-57;  
Matches 112; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
QY 1 VDSPIEQPRIIPNGGTLTLLGNAPKALRNEERAIDELKKQAIEDKEATTAIEAASSD 60  
DB 21 VDSPIEQPRIIPNGGTLTLLGNAPKALRNEERAIDELKKQAIEDKEATTAIEAASSD 80  
QY 61 ALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVQSDNAASDAWEK 120  
DB 81 ALEALADQADALQSEEAAYVQSDNAASDALEALADQADALQSEEAAYVVKADNAASDTLEA 140  
QY 121 AA 122  
DB 141 LA 142  
RESULT 9  
AAV71037  
ID AAV71037 standard; peptide; 58 AA.  
XX  
AC AAV71037;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 GRAB protein fragment #2.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 2; Page 55; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
related alpha2M binding protein) from Streptococcus pyogenes which have  
the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence



PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
PI WPI; 2000-365572/31.  
DR  
XX N-PSDB; AAD00562.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 5; Page 60-61; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a partial GRAB protein from S. pyogenes strain APl.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.  
XX  
SQ Sequence 155 AA;  
Query Match 96.5%; Score 750; DB 21; Length 155;  
Best Local Similarity 99.4%; Pred. No. 2e-84;  
Matches 154; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 2 DSPIEQPIIPNGGTLNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEASDDA 61  
DB 1 DSPIEQPIIPNGGTLNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEASDDA 60  
OY 62 LEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVDNAASDAWEKA 121  
DB 61 LEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVDNAASDAWEKA 120  
OY 122 ATPIALDVKKTKDKPKVVKKEERQNVNTLPTTGE 155  
DB 121 ATPIALDVKKTKDKPKVVKKEERQNVNTLPTTGE 155  
RESULT 6  
AAV71040  
ID AAV71040 standard; peptide; 141 AA.  
XX  
AC AAV71040;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 GRAB protein fragment #4.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX

PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 5; Page 57; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-174. This fragment is devoid of the trans-membrane  
CC and cell wall anchor regions. It is useful in vaccine composition.  
XX  
SQ Sequence 141 AA;  
Query Match 87.8%; Score 682; DB 21; Length 141;  
Best Local Similarity 100.0%; Pred. No. 4.4e-76;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VDSPIEQPIIPNGGTLNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEASDD 60  
DB 1 VDSPIEQPIIPNGGTLNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEASDD 60  
OY 61 ALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVDNAASDAWEK 120  
DB 61 ALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVDNAASDAWEK 120  
OY 121 AATPIALDVKKTKDKPKVVK 141  
DB 121 AATPIALDVKKTKDKPKVVK 141  
RESULT 7  
AAV71045  
ID AAV71045 standard; Protein; 271 AA.  
XX  
AC AAV71045;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain AP49 partial GRAB protein.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
DR N-PSDB; AAD00563.  
XX

FT Region 183..188  
/note="consensus sequence for gram-positive  
FT surface cell wall anchored proteins."  
FT 193..217  
FT /label=" Membrane\_spanning\_region  
PN WO200026240-A2.  
XX  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
DR N-PSDB; AAD00559, AAD00560.  
XX  
XX New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 5; Fig 2b; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a full-length GRAB protein from S. pyogenes strain Sf370.  
XX  
SQ Sequence 217 AA;  
Query Match 100.0%; Score 777; DB 21; Length 217;  
Best Local Similarity 100.0%; Pred. No. 1.4e-87;  
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDSPIEQPIIPNGGTLNLGNAPKALRNEERAIDELKKAIEDKEATTATIEAASSD 60  
DB 34 VDSPIEQPIIPNGGTLNLGNAPKALRNEERAIDELKKAIEDKEATTATIEAASSD 93  
QY 61 ALEALADQTDALQSEEAAYVKADNAASDALEALADQTDALQSEEAAYVQSDNAASDAWEK 120  
DB 94 ALEALADQTDALQSEEAAYVKADNAASDALEALADQTDALQSEEAAYVQSDNAASDAWEK 153  
QY 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEESNP 159  
DB 154 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEESNP 192  
RESULT 4  
AA71046  
ID AAY71046 standard; Protein; 167 AA.  
XX  
AC AAY71046;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain KTL3 partial GRAB protein.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 17  
FT /note="Thr at position 18 of GRAB protein fragment  
FT (AAY71036) from S. pyogenes strain Sf370 is replaced  
FT with Ile"  
XX  
PN WO200026240-A2.  
XX

PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
DR N-PSDB; AAD00564.  
XX  
XX New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 5; Page 62-63; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a partial GRAB protein from S. pyogenes strain KTL3.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.  
XX  
SQ Sequence 167 AA;  
Query Match 97.8%; Score 760; DB 21; Length 167;  
Best Local Similarity 100.0%; Pred. No. 1.3e-85;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDSPIEQPIIPNGGTLNLGNAPKALRNEERAIDELKKAIEDKEATTATIEAASSD 60  
DB 12 VDSPIEQPIIPNGGTLNLGNAPKALRNEERAIDELKKAIEDKEATTATIEAASSD 71  
QY 61 ALEALADQTDALQSEEAAYVKADNAASDALEALADQTDALQSEEAAYVQSDNAASDAWEK 120  
DB 72 ALEALADQTDALQSEEAAYVKADNAASDALEALADQTDALQSEEAAYVQSDNAASDAWEK 131  
QY 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGE 156  
DB 132 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGE 167  
RESULT 5  
AA71044  
ID AAY71044 standard; Protein; 155 AA.  
XX  
AC AAY71044;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain APl partial GRAB protein.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 17  
FT /note="Thr at position 18 of GRAB protein fragment  
FT (AAY71036) from S. pyogenes strain Sf370 is replaced  
FT with Ile"  
XX  
PN WO200026240-A2.  
XX

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-192 and devoid of the membrane spanning region. This  
CC fragment is useful in vaccine composition.

XX Sequence 159 AA;  
SQ Query Match 100.0%; Score 777; DB 21; Length 159;  
Best Local Similarity 100.0%; Pred. No. 9.9e-88;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VDSPIEQPRIIPNGGTLTLNLGNAPKALRNEERAIDELKQAIEDKATTAIEAASD 60  
DQ 1 VDSPIEQPRIIPNGGTLTLNLGNAPKALRNEERAIDELKQAIEDKATTAIEAASD 60  
OY 61 ALEALADQTDALQSEEAAYVVKADNASDALEALADQTDALQSEEAAYVVDNNAASDAWEK 120  
DQ 61 ALEALADQTDALQSEEAAYVVKADNASDALEALADQTDALQSEEAAYVVDNNAASDAWEK 120  
OY 121 AATPTALDVKKTKDKPVPVKKERQNVNTLPTTGESNP 159  
DQ 121 AATPTALDVKKTKDKPVPVKKERQNVNTLPTTGESNP 159

RESULT 2  
AAV71039  
ID AAY71039 standard; Protein: 184 AA.  
XX AC AAY71039;  
XX 29-AUG-2000 (first entry)  
XX Streptococcus pyogenes strain SF370 mature GRAB protein.  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
OS Streptococcus pyogenes.

FH Key Location/Qualifiers  
FT Binding-site 1..58  
FT Region /label= alpha2-macroglobulin\_binding\_site  
FT Region 59..86  
FT Region /label= R1  
FT Region /note= "repeat region"  
FT Region 87..114  
FT Region /label= R2  
FT Region /note= "repeat region"  
FT Region 115..159  
FT Region /label= Cell\_wall\_spanning\_region  
FT Region 150..155  
FT Region /note= "consensus sequence for gram-positive  
FT surface cell wall anchored proteins"  
FT Region 160..184  
FT Region /label= Membrane\_spanning\_region

XX WO200026240-A2.  
XX 11-MAY-2000.  
XX 02-NOV-1999; 99WO-GB03631.  
XX 02-NOV-1998; 98GB-0023975.  
XX

PA (ACTI-) ACTINOVA LTD.  
XX BJORCK LH, Rasmussen M;  
XX WPI: 2000-365572/31.  
XX New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX Claim 5; Page 56; 67pp; English.  
XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a mature GRAB protein from S. pyogenes strain SF370 without the  
CC signal sequence. This sequence is capable of binding alpha2M  
CC and useful in vaccine composition.

XX Sequence 184 AA;  
SQ Query Match 100.0%; Score 777; DB 21; Length 184;  
Best Local Similarity 100.0%; Pred. No. 1.2e-87;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VDSPIEQPRIIPNGGTLTLNLGNAPKALRNEERAIDELKQAIEDKATTAIEAASD 60  
DQ 1 VDSPIEQPRIIPNGGTLTLNLGNAPKALRNEERAIDELKQAIEDKATTAIEAASD 60  
OY 61 ALEALADQTDALQSEEAAYVVKADNASDALEALADQTDALQSEEAAYVVDNNAASDAWEK 120  
DQ 61 ALEALADQTDALQSEEAAYVVKADNASDALEALADQTDALQSEEAAYVVDNNAASDAWEK 120  
OY 121 AATPTALDVKKTKDKPVPVKKERQNVNTLPTTGESNP 159  
DQ 121 AATPTALDVKKTKDKPVPVKKERQNVNTLPTTGESNP 159

RESULT 3  
AAV71042  
ID AAY71042 standard; Protein: 217 AA.  
XX AC AAY71042;  
XX 29-AUG-2000 (first entry)  
XX Streptococcus pyogenes strain SF370 full-length GRAB protein.  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
OS Streptococcus pyogenes.  
FH Key Location/Qualifiers  
FT Peptide 1..33  
FT Protein /label= Signal\_sequence  
FT Protein 34..217  
FT Binding-site /label= Mature\_GRAB\_protein  
FT Binding-site 34..91  
FT Domain /label= alpha2M\_binding\_site  
FT Domain 34..68  
FT Region /note= "shows homology to E domain of protein G"  
FT Region 92..119  
FT Region /label= Repeat\_region\_1  
FT Region 120..147  
FT Region /label= Repeat\_region\_2  
FT Region 148..192  
FT Region /label= Cell\_wall\_spanning\_region

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 00:48:06 ; Search time 64.6203 Seconds  
(without alignments)  
273.300 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQPIIPNGTTLNL.....KKERQNVNTLPTTGESNP 159

Scoring table: BLOSUM62

Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	777	100.0	159	21	AAV71041 Streptococcus pyog
2	777	100.0	184	21	AAV71039 Streptococcus pyog
3	777	100.0	217	21	AAV71042 Streptococcus pyog
4	760	97.8	167	21	AAV71046 Streptococcus pyog
5	750	96.5	155	21	AAV71044 Streptococcus pyog
6	682	87.8	141	21	AAV71040 Streptococcus pyog
7	536	69.0	271	21	AAV71045 Streptococcus pyog
8	534	68.7	259	21	AAV71043 Streptococcus pyog
9	285	36.7	58	21	AAV71037 Streptococcus pyog
10	155	19.9	448	8	AAV70493 Protein G. Strept
11	155	19.9	448	10	AAV95030 Protein G. Strept

12	155	19.9	448	11	AAV70113 Protein G variant
13	155	19.9	448	12	AAV10004 Streptococcus Gx78
14	155	19.9	448	15	AAV53290 Streptococcus prot
15	155	19.9	593	11	AAV07014 Protein G variant
16	155	19.9	593	15	AAV62944 Streptococcus prot
17	155	19.9	594	12	AAV10005 Streptococcus Gx78
18	142	18.3	103	21	AAV57611 Streptococcus stra
19	129	16.6	28	21	AAV71038 Streptococcus pyog
20	120	15.4	23	21	AAV71036 Streptococcus pyog
21	115	14.8	25	21	AAV71048 Streptococcus pyog
22	105	13.5	235	10	AAV94785 Protein G variant.
23	105	13.5	235	11	AAV07004 Protein G variant.
24	105	13.5	265	12	AAV10011 Type 4 Gx7809 prot
25	105	13.5	265	15	AAV53294 IgG-binding Strept
26	97	12.5	20	21	AAV71051 Streptococcus pyog
27	96	12.4	19	21	AAV71047 Streptococcus pyog
28	93	12.0	422	22	ABG16038 Novel human diagno
29	93	12.0	459	22	ABG20309 Novel human diagno
30	91	11.7	413	16	AAV71928 S. dysgalactiae MA
31	91	11.7	413	16	AAV71670 S. dysgalactiae MA
32	90	11.6	480	8	AAV70468 Sequence of polype
33	86	11.1	19	21	AAV71049 Streptococcus pyog
34	86	11.1	1107	22	ABG18247 Novel human diagno
35	85	10.9	664	16	AAV71929 S. dysgalactiae MI
36	84	10.8	489	22	ABV58655 Drosophila melanog
37	83	10.7	46	16	AAV71125 SG-3. Synthetic.
38	83	10.7	168	21	AAV20163 Arabidopsis thalia
39	83	10.7	180	21	AAV20162 Arabidopsis thalia
40	83	10.7	200	16	AAV71128 Synthetic protein
41	83	10.7	208	16	AAV71127 Synthetic protein
42	82	10.6	18	21	AAV71050 Streptococcus pyog
43	81	10.4	1022	22	ABG25650 Novel human diagno
44	76	9.8	3248	17	AAV99795 Kinetochore protel
45	75	9.7	429	22	AAV01027 CFE 30 protein seq

#### ALIGNMENTS

RESULT 1

AAV71041

ID AAV71041 standard; peptide: 159 AA.

AC AAV71041;

DT 29-AUG-2000 (first entry)

XX Streptococcus pyogenes strain SF370 GRAB protein fragment #5.

DE GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

KW Immune response; Streptococcus pyogenes infection.

OS Streptococcus pyogenes.

XX Streptococcus pyogenes.

PN WO200026240-42.

XX 11-MAY-2000.

PD 11-MAY-2000.

PF 02-NOV-1999 99WO-GB03631.

PR 02-NOV-1998; 98GB-0023975.

XX (ACTI-) ACTINOVA LTD.

PA Bjorck LH, Rasmussen M;

```
OS Balanus amphitrite.
XX
FH Key Location/Qualifiers
FT CDS 1..1371
FT /*tag= a
FT /product= sixth_adhesion_protein
XX
XX JP11332573-A.
XX
XX 07-DEC-1999.
XX
XX 29-MAY-1998; 98JP-0149138.
XX
XX 29-MAY-1998; 98JP-0149138.
XX
XX (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
XX
XX WPI; 2000-091356/08.
XX P-PSDB; AAY67238.
XX
XX Barnacle sixth adhesion protein gene - useful as a raw material for
PT adhesives
XX
XX Example 4; Page 4-5; 8pp; Japanese.
XX
XX This is the coding sequence of the barnacle sixth adhesion protein. The
CC protein was isolated from the cement secreted from the shell of
CC barnacles. The protein can be used as an adhesive in water, and in wet
CC environments.
XX
XX Sequence 1773 BP; 429 A; 499 C; 524 G; 321 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 333 Length: 1773
XX Score: 45.00 Matches: 11
XX Percent Similarity: 53.85% Conservative: 3
XX Best Local Similarity: 42.31% Mismatches: 12
XX Query Match: 34.88% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-847-539A-6_COPY_59_86 (1-28) x AA256162 (1-1773)
XX
XX Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
XX ||||||| : : : : :
XX Db 599 GCATTGGAGACTGTGCTGGATCCTCCGGGAACGTCACGTCCTCGAGCGCGGTGTTG 540
XX
XX Qy 23 LysAlaAspAsnAlaAla 28
XX |||
XX Db 539 AGCGCCTCAGAAAGCTGCG 522
XX
XX Search completed: October 13, 2002, 02:41:30
XX Job time : 1028.43 secs
```



```
PD 02-DEC-1999.
XX
XX
PF 21-MAY-1999; 99WO-AU00385.
XX
XX
PR 21-MAY-1998; 98AU-0003634.
XX
XX
PA (UNSY ) UNIV SYDNEY.
XX
XX
PI Reeves PR, Wang L;
XX
XX
DR WPI; 2000-072598/06.
XX
XX
PT Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, faeces, etc.
XX
XX
PS Claim 3; Page 200-201; 245pp; English.
XX
XX
CC AA256331 to AA256398 represent nucleic acid molecules (I) encoding all
CC or part of an Escherichia coli flagellin protein except a protein
CC expressed by E. coli H1, H7, H12 or H48 type strains. The present
CC invention also describes a method of detecting the presence of E. coli
CC of a particular H serotype in a sample, comprising specifically
CC hybridising a nucleic acid, preferably at least a pair, derived from a
CC flagellating gene, specific for a particular flagellin gene associated
CC with the H serotype, to any E. coli in the sample which contain the gene,
CC and detecting any hybridised molecules, identifying the presence of that
CC serotype in the sample. (I) are useful for: (1) detecting the presence
CC of E. coli of H serotype in a sample by hybridising at least one or a
CC pair of (I) to any E. coli in the sample and detecting the hybridised
CC nucleic acid molecules; and (2) for detecting the presence of both O
CC and H-serotypes of E. coli by hybridising at least one or a pair of (I)
CC to any E. coli present in the sample and detecting the hybridised
CC nucleic acid molecules. (I) is particularly useful for detecting the
CC combination of O and H antigen. Hybridised (I) when using at least one
CC (I) is detected by southern blot analysis and, when using a pair of (I),
CC is detected by polymerase chain reaction (PCR). AA256399 to AA256420
CC represent primers used in the exemplification of the present invention.
XX
XX
SQ Sequence 1380 BP; 424 A; 299 C; 309 G; 348 T; 0 other;

Alignment Scores:
Pred. No.: 243 Length: 1380
Score: 45.00 Matches: 9
Percent Similarity: 64.00% Conservative: 7
Best Local Similarity: 36.00% Mismatches: 9
Query Match: 34.88% Indels: 0
DB: 21 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AA256351 (1-1380)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
DB 1123 GATGCAGCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAGACTATCGACAAGCATTTG 1182
QY 22 ValLysAlaAspAsn 26
DB 1183 GCTAAAGTTGACAAT 1197

RESULT 37
AAZ56376
ID AA256376 standard; DNA; 1380 BP.
XX
XX
AC AA256376;
XX
XX
DT 17-MAR-2000 (first entry)
XX
XX
DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:46.
XX
XX
KW Flagellin; fliC; antigen; detection; ds.
XX
XX
OS Escherichia coli.
XX
PN WO961458-A1.
```

```
XX
PD 02-DEC-1999.
XX
XX
PF 21-MAY-1999; 99WO-AU00385.
XX
XX
PR 21-MAY-1998; 98AU-0003634.
XX
XX
PA (UNSY ) UNIV SYDNEY.
XX
XX
PI Reeves PR, Wang L;
XX
XX
DR WPI; 2000-072598/06.
XX
XX
PT Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, faeces, etc.
XX
XX
PS Claim 3; Page 22; 245pp; English.
XX
XX
CC AA256331 to AA256398 represent nucleic acid molecules (I) encoding all
CC or part of an Escherichia coli flagellin protein except a protein
CC expressed by E. coli H1, H7, H12 or H48 type strains. The present
CC invention also describes a method of detecting the presence of E. coli
CC of a particular H serotype in a sample, comprising specifically
CC hybridising a nucleic acid, preferably at least a pair, derived from a
CC flagellating gene, specific for a particular flagellin gene associated
CC with the H serotype, to any E. coli in the sample which contain the gene,
CC and detecting any hybridised molecules, identifying the presence of that
CC serotype in the sample. (I) are useful for: (1) detecting the presence
CC of E. coli of H serotype in a sample by hybridising at least one or a
CC pair of (I) to any E. coli in the sample and detecting the hybridised
CC nucleic acid molecules; and (2) for detecting the presence of both O
CC and H-serotypes of E. coli by hybridising at least one or a pair of (I)
CC to any E. coli present in the sample and detecting the hybridised
CC nucleic acid molecules. (I) is particularly useful for detecting the
CC combination of O and H antigen. Hybridised (I) when using at least one
CC (I) is detected by southern blot analysis and, when using a pair of (I),
CC is detected by polymerase chain reaction (PCR). AA256399 to AA256420
CC represent primers used in the exemplification of the present invention.
XX
XX
SQ Sequence 1380 BP; 419 A; 299 C; 313 G; 349 T; 0 other;

Alignment Scores:
Pred. No.: 243 Length: 1380
Score: 45.00 Matches: 9
Percent Similarity: 64.00% Conservative: 7
Best Local Similarity: 36.00% Mismatches: 9
Query Match: 34.88% Indels: 0
DB: 21 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AA256376 (1-1380)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
DB 1123 GATGCAGCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAGACTATCGACAAGCATTTG 1182
QY 22 ValLysAlaAspAsn 26
DB 1183 GCTAAAGTTGACAAT 1197

RESULT 38
AAZ56340
ID AA256340 standard; DNA; 1383 BP.
XX
XX
AC AA256340;
XX
XX
DT 17-MAR-2000 (first entry)
XX
XX
DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:10.
XX
XX
KW Flagellin; fliC; antigen; detection; ds.
XX
XX
OS Escherichia coli.
XX
```

```

PF 21-MAY-1999; 99WO-AU00385.
XX
PR 21-MAY-1998; 98AU-0003634.
XX
XX (UNSY ) UNIV SYDNEY.
XX
PI Reeves PR, Wang L;
XX
DR WPI; 2000-072598/06.
XX
XX Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, faeces, etc.
XX
PS Claim 3; Page 204; 245pp; English.
XX
CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all
CC or part of an Escherichia coli flagellin protein except a protein
CC expressed by E. coli H1, H7, H12 or H48 type strains. The present
CC invention also describes a method of detecting the presence of E. coli
CC of a particular H serotype in a sample, comprising specifically
CC hybridising a nucleic acid, preferably at least a pair, derived from a
CC flagellating gene, specific for a particular flagellin gene associated
CC with the H serotype, to any E. coli in the sample which contain the gene,
CC and detecting any hybridised molecules, identifying the presence of that
CC serotype in the sample. (I) are useful for: (1) detecting the presence
CC of E. coli of H serotype in a sample by hybridising at least one or a
CC pair of (I) to any E. coli in the sample and detecting the hybridised
CC nucleic acid molecules; and (2) for detecting the presence of both O
CC and H-serotypes of E. coli by hybridising at least one or a pair of (I)
CC to any E. coli present in the sample and detecting the hybridised
CC nucleic acid molecules. (I) is particularly useful for detecting the
CC combination of O and H antigen. Hybridised (I) when using at least one
CC (I) is detected by southern blot analysis and, when using a pair of (I),
CC is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420
CC represent primers used in the exemplification of the present invention.
XX
SQ Sequence 1365 BP; 412 A; 289 C; 312 G; 352 T; 0 other;

Alignment Scores:
Pred. No.: 239 Length: 1365
Score: 45.00 Matches: 9
Percent Similarity: 64.00% Conservative: 7
Best Local Similarity: 36.00% Mismatches: 9
Query Match: 34.88% Indels: 0
DB: 21 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AAZ56356 (1-1365)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
Db 1108 GATGCAGCAAAATCGTTGCAATCTACTACCAACCCGCTCGAAGCATTG 1167
QY 22 ValLysAlaAspAsn 26
Db 1168 GCTAAAGTTGACAAT 1182

RESULT 35
AAZ56343
ID AAZ56343 standard; DNA; 1368 BP.
XX
AC AAZ56343;
XX
XX 17-MAR-2000 (first entry)
XX
DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:13.
XX
KW Flagellin; fliC; antigen; detection; ds.
XX
OS Escherichia coli.
XX
PN WO9961458-A1.
XX
PD 02-DEC-1999.

```

```

XX
PF 21-MAY-1999; 99WO-AU00385.
XX
PR 21-MAY-1998; 98AU-0003634.
XX
XX (UNSY ) UNIV SYDNEY.
XX
PI Reeves PR, Wang L;
XX
DR WPI; 2000-072598/06.
XX
XX Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, faeces, etc.
XX
PS Claim 3; Page 195; 245pp; English.
XX
CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all
CC or part of an Escherichia coli flagellin protein except a protein
CC expressed by E. coli H1, H7, H12 or H48 type strains. The present
CC invention also describes a method of detecting the presence of E. coli
CC of a particular H serotype in a sample, comprising specifically
CC hybridising a nucleic acid, preferably at least a pair, derived from a
CC flagellating gene, specific for a particular flagellin gene associated
CC with the H serotype, to any E. coli in the sample which contain the gene,
CC and detecting any hybridised molecules, identifying the presence of that
CC serotype in the sample. (I) are useful for: (1) detecting the presence
CC of E. coli of H serotype in a sample by hybridising at least one or a
CC pair of (I) to any E. coli in the sample and detecting the hybridised
CC nucleic acid molecules; and (2) for detecting the presence of both O
CC and H-serotypes of E. coli by hybridising at least one or a pair of (I)
CC to any E. coli present in the sample and detecting the hybridised
CC nucleic acid molecules. (I) is particularly useful for detecting the
CC combination of O and H antigen. Hybridised (I) when using at least one
CC (I) is detected by southern blot analysis and, when using a pair of (I),
CC is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420
CC represent primers used in the exemplification of the present invention.
XX
SQ Sequence 1368 BP; 419 A; 292 C; 312 G; 345 T; 0 other;

Alignment Scores:
Pred. No.: 240 Length: 1368
Score: 45.00 Matches: 9
Percent Similarity: 64.00% Conservative: 7
Best Local Similarity: 36.00% Mismatches: 9
Query Match: 34.88% Indels: 0
DB: 21 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AAZ56343 (1-1368)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
Db 1111 GATGCAGCAAAATCGTTGCAATCTACTACCAACCCGCTCGAAGCATTG 1170
QY 22 ValLysAlaAspAsn 26
Db 1171 GCTAAAGTTGACAAT 1185

RESULT 36
AAZ56351
ID AAZ56351 standard; DNA; 1380 BP.
XX
AC AAZ56351;
XX
XX 17-MAR-2000 (first entry)
XX
DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:21.
XX
KW Flagellin; fliC; antigen; detection; ds.
XX
OS Escherichia coli.
XX
PN WO9961458-A1.
XX

```



XX Claim 3; Page 190; 245pp; English.  
 XX AAZ56331 to AAZ56398 represent nucleic acid molecules (1) encoding all  
 CC or part of an *Escherichia coli* flagellin protein except a protein  
 CC expressed by *E. coli* H1, H7, H12 or H48 type strains. The present  
 CC invention also describes a method of detecting the presence of *E. coli*  
 CC of a particular H serotype in a sample, comprising specifically  
 CC hybridising a nucleic acid, preferably at least a pair, derived from a  
 CC flagellating gene, specific for a particular flagellin gene associated  
 CC with the H serotype, to any *E. coli* in the sample which contain the gene,  
 CC and detecting any hybridised molecules, identifying the presence of that  
 CC serotype in the sample. (1) are useful for: (1) detecting the presence  
 CC of *E. coli* of H serotype in a sample by hybridising at least one or a  
 CC pair of (1) to any *E. coli* in the sample and detecting the hybridised  
 CC nucleic acid molecules; and (2) for detecting the presence of both O  
 CC and H-serotypes of *E. coli* by hybridising at least one or a pair of (1)  
 CC to any *E. coli* present in the sample and detecting the hybridised  
 CC nucleic acid molecules. (1) is particularly useful for detecting the  
 CC combination of O and H antigen. Hybridised (1) when using at least one  
 CC (1) is detected by southern blot analysis and, when using a pair of (1),  
 CC is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420  
 CC represent primers used in the exemplification of the present invention.  
 XX SQ Sequence 417 BP; 120 A; 108 C; 95 G; 94 T; 0 other;

Alignment Scores:  
 Pred. No.: 53.7 Length: 417  
 Score: 45.00 Matches: 9  
 Percent Similarity: 64.00% Conservative: 7  
 Best Local Similarity: 36.00% Mismatches: 9  
 Query Match: 34.88% Indels: 0  
 DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAZ56335 (1-417)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21  
 Db 135 GATGACGAAATCGTTCGAATCTACCAACCGCTCGAAACTATCGCAAGCATTTG 194

QY 22 VallysAlaAspAsn 26  
 Db 195 GCTAAGTTGACAAAT 209

RESULT 33

AAF08476/c

ID AAF08476 standard; cDNA; 461 BP.

AC AAF08476;

DT 13-MAR-2001 (first entry)

DE Fusarium venenatum EST SEQ ID NO:999.

XX Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; *Fusarium venenatum*; *Aspergillus niger*;  
 KW *Aspergillus oryzae*; *Trichoderma reesei*; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.

OS Fusarium venenatum.

XX WO200056762-A2.

PW 28-SEP-2000.

XX 22-MAR-2000; 2000WO-US07781.

XX 22-MAR-1999; 99US-0273623.

XX (NOVO ) NOVO NORDISK BIOTECH INC.

PA (NOVO ) NOVO NORDISK AS.

XX

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 XX WPI; 2000-594572/56.  
 XX Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -  
 XX Claim 86; Page 763-764; 3161pp; English.

XX The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from *Aspergillus*  
 CC *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and  
 CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are  
 CC all specifically claimed in the present invention.

XX SQ Sequence 461 BP; 117 A; 131 C; 122 G; 89 T; 2 other;

Alignment Scores:  
 Pred. No.: 61 Length: 461  
 Score: 45.00 Matches: 10  
 Percent Similarity: 62.50% Conservative: 5  
 Best Local Similarity: 41.67% Mismatches: 9  
 Query Match: 34.88% Indels: 0  
 DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAF08476 (1-461)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallysAla 24

Db 176 GAAGCACTCACGGCACAGGTTCGACTTTCGATAAGACCGGCTTGCTTGCATGCA 117

QY 25 AspAsnAlaAla 28

Db 116 GACACGGCAGCT 105

RESULT 34

AAZ56356

ID AAZ56356 standard; DNA; 1365 BP.

XX AAZ56356;

AC AAZ56356;

DT 17-MAR-2000 (first entry)

DE *Escherichia coli* flagellin protein nucleotide sequence SEQ ID NO:26.

XX Flagellin; *flhC*; antigen; detection; ds.

OS *Escherichia coli*.

PN WO9961458-A1.

XX 02-DEC-1999.

XX

CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a nucleic acid described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

XX SQ Sequence 349980 BP; 80289 A; 91081 C; 97378 G; 81232 T; 0 other;

Alignment Scores:  
 Pred. No.: 1.69e+05 Length: 349980  
 Score: 46.00 Matches: 9  
 Percent Similarity: 62.50% Conservative: 6  
 Best Local Similarity: 37.50% Mismatches: 9  
 Query Match: 35.66% Indels: 0  
 DB: 22 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAH68525 (1-349980)

OY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
 DB 225435 ATTGAGCCTTACGCACCTGAGCGAGCAGCAGCTTGAGAGGCGCGAGTCCGCAATCTACAAG 225494  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: |||

OY 24 AlaAspAsnAla 27

DB 225495 TTCGAGAAGCGCT 225506  
 :::: ||| |||

RESULT 31

AA81490  
 ID AAA81490 standard; DNA; 1437668 BP.

XX AC AAA81490;

DT 04-DEC-2000 (first entry)

DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX Neisseria meningitidis.

XX WO200022430-A2.

PD 20-APR-2000.

PF 08-OCT-1999; 99WO-US23573.

PR 09-OCT-1998; 98US-0103794.

PR 30-APR-1999; 99US-0132068.

PA (CHIR ) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;  
 PI Rappuoli R, Pizza M;  
 XX WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
 PT used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N.gonorrhoea -

PS Claim 7; Page 866-1272; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA

CC sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins: AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

XX SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

Alignment Scores:  
 Pred. No.: 8.9e+05 Length: 1437668  
 Score: 46.00 Matches: 11  
 Percent Similarity: 65.38% Conservative: 6  
 Best Local Similarity: 42.31% Mismatches: 9  
 Query Match: 35.66% Indels: 0  
 DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAA81490 (1-1437668)

OY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22  
 :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: |||

DB 411891 GCTTTGAGAGCGCGCTTTGGAAAACTCGAGCCCGCAAAATCCGACGGCAGCATTTG 411950

OY 23 LysAlaAspAsnAlaAla 28

DB 411951 GAAGCGGCTGAAGCGGCT 411968  
 :::: ||| |||

RESULT 32

AAZ56335  
 ID AAZ56335 standard; DNA; 417 BP.

XX AC AAZ56335;

DT 17-MAR-2000 (first entry)

DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:5.

XX Flagellin; fliC; antigen; detection; ds.

XX Escherichia coli.

XX WO9961458-A1.

PD 02-DEC-1999.

PF 21-MAY-1999; 99WO-AU00385.

PR 21-MAY-1998; 98AU-0003634.

PA (UNSY ) UNIV SYDNEY.

XX Reeves PR, Wang L;

XX WPI; 2000-072598/06.

XX Novel nucleic acid molecule useful for the detection of flagellated  
 PT bacterial strains in food, faeces, etc. -

```
XX
SQ Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other:

Alignment Scores:
Pred. No.: 2.28e+04 Length: 69936
Score: 46.00 Matches: 11
Percent Similarity: 65.38% Conservative: 6
Best Local Similarity: 42.31% Mismatches: 9
Query Match: 35.66% Indels: 0
DB: 21 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AAA81479 (1-69936)
Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
Db 69296 GCTTTGGAGCGCGTTTGGAAAACTCGAAGCCGCGCAAAATCCGACGGCAGCATTTG 69237

Qy 23 LysAlaAspAsnAlaAla 28
Db 69236 GAAGCGGCTGAAGCGGCT 69219

RESULT 29
AAF21607
ID AAF21607 standard; DNA; 349980 BP.
AC AAF21607;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
OS Neisseria meningitidis.
XX
PN WO200056791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
DR WPI; 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections .
XX
PS Claim 7; Appendix A; 692pp; English.
XX

The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
```

```
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
SQ Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other:

Alignment Scores:
Pred. No.: 1.69e+05 Length: 349980
Score: 46.00 Matches: 11
Percent Similarity: 65.38% Conservative: 6
Best Local Similarity: 42.31% Mismatches: 9
Query Match: 35.66% Indels: 0
DB: 21 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AAF21607 (1-349980)
Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
Db 111891 GCTTTGGAGCGCGTTTGGAAAACTCGAAGCCGCGCAAAATCCGACGGCAGCATTTG 111950

Qy 23 LysAlaAspAsnAlaAla 28
Db 111951 GAAGCGGCTGAAGCGGCT 111968

RESULT 30
AAH68525
ID AAH68525 standard; DNA; 349980 BP.
XX
AC AAH68525;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 7060.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Havashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
PS Disclosure; SEQ ID NO: 7060; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
```

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
:::||||| ||| :::: ||| |||::: |||::: |||::: |||  
Db 1540 ATTGAGCCTTACGACCTGAGGAGGAGGAGCTTGAAGAGGGCGAGTCCGCAATCTACAAG 1599

QY 24 AlaAspAsnAla 27  
:::|||||  
Db 1600 TTCGAGAACGCT 1611

RESULT 27  
AAH65521/C  
ID AAH65521 standard; DNA; 2160 BP.  
XX  
AC AAH65521;  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum coding sequence fragment SEQ ID NO: 556.  
XX  
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-0127688.  
XX  
PR 16-DEC-1999; 99JP-0377484.  
XX  
PR 07-APR-2000; 2000JP-0159162.  
XX  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
WPI: 2001-376931/40.  
DR P-PSDB; AAG90302.  
XX

Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene

Claim 8: SEQ ID NO: 556; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Coryneform bacterium, and identifying a homolog of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Sequence 2160 BP; 462 A; 569 C; 621 G; 508 T; 0 other;

Alignment Scores:  
Pred. No.: 286 Length: 2160  
Score: 46.00 Matches: 9  
Percent Similarity: 62.50% Conservative: 6  
Best Local Similarity: 37.50% Mismatches: 9  
Query Match: 35.66% Indels: 0  
DB: 22 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAH65521 (1-2160)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
:::||||| ||| :::: ||| |||::: |||::: |||::: |||  
Db 636 ATTGAGCCTTACGACCTGAGGAGGAGGAGCTTGAAGAGGGCGAGTCCGCAATCTACAAG 577

QY 24 AlaAspAsnAla 27  
:::|||||  
Db 576 TTCGAGAACGCT 565

RESULT 28  
AAH81479/C  
ID AAH81479 standard; DNA; 69936 BP.  
XX  
AC AAH81479;  
DT 04-DEC-2000 (first entry)  
XX  
DE N. meningitidis partial DNA sequence gnm\_27 SEQ ID NO: 27.  
XX  
DE Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
KW Meningococcus B; MenB; ds.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO200022430-A2.  
XX  
PD 20-APR-2000.  
XX  
PF 08-OCT-1999; 99WO-US23573.  
XX  
PR 09-OCT-1998; 98US-0103794.  
XX  
PR 30-APR-1999; 99US-0132068.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
PI Rappuoli R, Pizza M;  
XX  
WPI: 2000-318079/27.  
XX

Isolated nucleotide sequences of Neisseria meningitidis which can be used in the diagnosis and treatment of N. meningitidis infection and other Neisserial infections, for example, N.gonorrhoea

Claim 7: Page 547-567; 1760pp; English.

The present invention describes methods of obtaining immunogenic proteins from Neisseria genomic sequences. AAA81453 to AAA82414 represent specifically claimed Neisseria meningitidis genomic DNA sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent Neisseria DNA sequences and their corresponding proteins; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to Neisserial bacteria. For example, some of the identified proteins could be components of vaccines against Meningococcus B; against all serotypes; and/or against all pathogenic Neisseriae. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious Meningococcus B vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

AAFI1343;  
 13-MAR-2001 (first entry)  
 Aspergillus niger EST SEQ ID NO:3866.  
 Multiple gene expression: filamentous fungal cell; EST;  
 expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 culture condition; environmental stress; spore morphogenesis;  
 metabolic pathway engineering; catabolic pathway engineering; ss.  
 Aspergillus niger.  
 WO200056762-A2.  
 28-SEP-2000.  
 22-MAR-2000; 2000WO-US07781.  
 22-MAR-1999; 99US-0273623.  
 (NOVO ) NOVO NORDISK BIOTECH INC.  
 (NOVO ) NOVO NORDISK AS.  
 Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 WPI; 2000-594572/56.  
 Monitoring differential expression of genes in filamentous fungal cells  
 uses fluorescence-labeled nucleic acids isolated from the cells and a  
 substrate of expressed sequence tags -  
 Claim 87; Page 1728; 3161pp; English.  
 The present invention describes a method for monitoring differential  
 expression of genes in a first filamentous fungal (FF) cell relative to  
 expression of the same genes in one or more second filamentous fungal  
 cells. The method uses fluorescence-labeled nucleic acids isolated from  
 the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 are used in the methods for monitoring differential expression of genes  
 in a first filamentous fungal (FF) cell relative to expression of the  
 same genes in one or more second filamentous fungal cells. Monitoring  
 the global expression of genes from FF cells allows the production  
 potential of the microorganisms to be improved. New genes may be  
 discovered, possible functions of unknown open reading frames can be  
 identified and gene copy number variation and stability can be  
 monitored. The expression of genes can be used to study how FF cells  
 adapt to changes in culture conditions, environmental stress, spore  
 morphogenesis, recombination, metabolic or catabolic pathway  
 engineering. Using ESTs provides several advantages over genomic or  
 random cDNA clones including elimination of redundancy as one spot on an  
 array equals one gene or open reading frame, and organisation of the  
 microarrays based on function of the gene products to facilitate  
 analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from  
 Aspergillus niger; AAF11854 to AAF14878 represents ESTs from Aspergillus  
 oryzae; and AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 all specifically claimed in the present invention.  
 Sequence 636 BP; 169 A; 155 C; 161 G; 149 T; 2 other;  
 Alignment Scores:  
 Pred. No.: 61.2 Length: 636  
 Score: 46.00 Matches: 11  
 Percent Similarity: 64.00% Conservative: 5  
 Best Local Similarity: 44.00% Mismatches: 9  
 Query Match: 35.66% Indels: 0  
 DB: 21 Gaps: 0  
 US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAF11343 (1-636)  
 Oy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20

Db 149 TCTAAAGCTGCAGAGCCCTTCCGAGAGATCGGCCCGCAAGTAATGTCAAGTCAGCT 208  
 Oy 21 ValValtysAlaAsp 25  
 Db 209 GCCTATATAAGCTGAT 223  
 RESULT 26  
 AAH65520  
 ID AAH65520 standard; DNA; 2115 BP.  
 XX  
 AC AAH65520;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum coding sequence fragment SEQ ID NO: 555.  
 XX  
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis; ds.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EP1108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX  
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 DR WPI; 2001-376931/40.  
 DR P-PSDB; AAG90301.  
 XX  
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 mutation point of a gene, measuring expression of a gene, analysing  
 expression profile or pattern of a gene and identifying homologous gene  
 -  
 Claim 8; SEQ ID NO: 555; 246pp + Sequence Listing; English.  
 XX  
 PS The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a nucleic acid described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 SQ Sequence 2115 BP; 492 A; 609 C; 561 G; 453 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 278 Length: 2115  
 Score: 46.00 Matches: 9  
 Percent Similarity: 62.50% Conservative: 6  
 Best Local Similarity: 37.50% Mismatches: 9  
 Query Match: 35.66% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAH65520 (1-2115)

```
Db 1897 GCTGCTGATGCGCTTGCA 1880
RESULT 23
AAI99683/c
ID AAI99683 standard; DNA; 4403765 BP.
XX
AC AAI99683;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
WPI; 2001-647261/74.
XX
Evaluating strain variation of Mycobacterium tuberculosis, comprises
determining the nucleotide sequence of the strain at positions in the
genome corresponding to positions where M. tuberculosis strains CDC
1551 and H37Rv differ -
XX
Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
The invention relates to evaluating strain variation within and between
different populations of the tuberculosis bacterial pathogen,
Mycobacterium tuberculosis or related Mycobacterium by determining the
nucleotide sequence of the first strain at positions in the complete
sequence of the genome that correspond to positions that differ in the
nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
H37Rv (AAI99682). The method is useful for evaluating strain variation of
M. tuberculosis and has valuable application in the fields of
tuberculosis genetics, epidemiology, patient treatment and epidemic
monitoring.
XX
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from USPTO
at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Alignment Scores:
Pred. No.: 1.93e+06 Length: 4403765
Score: 47.00 Matches: 9
Percent Similarity: 54.55% Conservative: 3
Best Local Similarity: 40.91% Mismatches: 10
Query Match: 36.43% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AAI99683 (1-4403765)
Qy 7 LeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValysAlaAspAsn 26
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4036256 TTGATGGACGCCGCGACCTGCTGGAGCGCAACAGGTAACCATGTCGTATCGACAAC 4036197
Qy 27 AlaAla 28
|||
Db 4036196 GGTGCT 4036191
|||
RESULT 24
```

```
AAI99682/c
ID AAI99682 standard; DNA; 4411529 BP.
XX
AC AAI99682;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
WPI; 2001-647261/74.
XX
Evaluating strain variation of Mycobacterium tuberculosis, comprises
determining the nucleotide sequence of the strain at positions in the
genome corresponding to positions where M. tuberculosis strains CDC
1551 and H37Rv differ -
XX
Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
The invention relates to evaluating strain variation within and between
different populations of the tuberculosis bacterial pathogen,
Mycobacterium tuberculosis or related Mycobacterium by determining the
nucleotide sequence of the first strain at positions in the complete
sequence of the genome that correspond to positions that differ in the
nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
H37Rv (AAI99682). The method is useful for evaluating strain variation of
M. tuberculosis and has valuable application in the fields of
tuberculosis genetics, epidemiology, patient treatment and epidemic
monitoring.
XX
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from USPTO
at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

Alignment Scores:
Pred. No.: 1.93e+06 Length: 4411529
Score: 47.00 Matches: 9
Percent Similarity: 54.55% Conservative: 3
Best Local Similarity: 40.91% Mismatches: 10
Query Match: 36.43% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AAI99682 (1-4411529)
Qy 7 LeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValysAlaAspAsn 26
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4044185 TTGATGGACGCCGCGACCTGCTGGAGCGCAACAGGTAACCATGTCGTATCGACAAC 4044126
Qy 27 AlaAla 28
|||
Db 4044125 GGTGCT 4044120
|||
RESULT 25
AAFI1343
ID AAFI1343 standard; CDNA; 636 BP.
XX
```

```
DB: 23 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AAS94455 (1-3711)
Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
    |||||
Db 1741 GCAGTGATCCAACTGACGCCGACAGCAGCGGTGGCGAGTGAAGACGCTCCGGTTGCA 1800

Qy 23 LysAlaAspAsn 26
    |||||
Db 1801 CCGCGCGACAAAT 1812

RESULT 21
ABL25849/c
ID ABL25849 standard; DNA: 1616 BP.
XX
AC ABL25849;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 29020.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 29020; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1616 BP; 457 A; 447 C; 439 G; 273 T; 0 other;

Alignment Scores:
Pred. No.: 88.8 Length: 1616
Score: 48.00 Matches: 12
Percent Similarity: 50.00% Conservative: 1
Best Local Similarity: 46.15% Mismatches: 13
Query Match: 37.21% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL25849 (1-1616)
Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
    |||||
```

```
Db 885 GCGGCTGATCTGCTGGTGGATGTTGGATGCGCTGGTTGGGAGCATGCTGCTGCTGCT 826
    ||| :||| ||| ||||| ||| |||||
Qy 23 LysAlaAspAsnAlaAla 28
    ||||| |||
Db 825 GCTGCTGATGCTGCTGCA 808
    ||||| |||

RESULT 22
ABL25848/c
ID ABL25848 standard; DNA: 4229 BP.
XX
AC ABL25848;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 29017.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 29017; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4229 BP; 1287 A; 1022 C; 928 G; 992 T; 0 other;

Alignment Scores:
Pred. No.: 299 Length: 4229
Score: 48.00 Matches: 12
Percent Similarity: 50.00% Conservative: 1
Best Local Similarity: 46.15% Mismatches: 13
Query Match: 37.21% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL25848 (1-4229)
Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
    ||||| ||| ||| ||||| ||| |||||
Db 1957 GCGGCTGATGCTGCTGGATGCTGGATGCTGGTGGGAGCATGCTGCTGCTGCT 1898
    ||||| ||| ||| ||||| ||| |||||
Qy 23 LysAlaAspAsnAlaAla 28
    ||||| |||
```

IDA03041 standard; DNA; 2944528 BP.  
ABA03041;  
DT 05-FEB-2002 (first entry)  
XX  
DE Listeria monocytogenes EGD-e genome sequence.  
XX  
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
vitamin B12; bacterial infection; disease; ds.  
XX  
OS Listeria monocytogenes.  
XX  
PN WO200177335-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-FR01118.  
XX  
PR 11-APR-2000; 2000FR-0004629.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Psihi H, Dehoux P;  
Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
P Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
Madenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
Rose M, Voss H;  
XX  
WPI: 2002-010914/01.  
DR  
XX  
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
and prevention of Listeria and related bacterial infections, and  
related polypeptides -  
PT  
PS Claim 1: SEQ ID No 1: 192pp; French.  
XX  
XX The present sequence is the genome sequence of Listeria monocytogenes  
EGD-e. This sequence and fragments of this sequence are useful for  
selecting probes and primers for detecting genes in L. monocytogenes and  
related organisms, and to study genetic polymorphisms and other genomes.  
CC  
CC Proteins (ABBA7237-ABB50149) expressed from the present sequence are  
useful for raising specific antibodies, identification of L.  
CC  
CC monocytogenes and related organisms, and for biosynthesis and  
biodegradation, especially biosynthesis of Vitamin B12. This sequence and  
proteins encoded by it are also useful for selecting compounds that  
regulate gene expression and cell replication and modulate L.  
CC  
CC monocytogenes-related diseases. In addition, this sequence and proteins  
encoded by it are useful in pharmaceutical and vaccines compositions for  
the treatment or prevention of infections by L. monocytogenes and related  
organisms.  
CC  
CC Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;  
Alignment Scores:  
Pred. No.: 3.25e+05 Length: 2944528  
Score: 51.00 Matches: 10  
Percent Similarity: 65.38% Conservative: 7  
Best Local Similarity: 38.46% Mismatches: 9  
Query Match: 39.53% Indels: 0  
DB: 24 Gaps: 0  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x ABA03041 (1-2944528)  
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 20  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||  
Db 51231 GCTGAACAAAAGAACAACTAGTATGATGAACAGATCGACTTGCACGAGATGCAGCT 51290

QY 21 ValValLysAlaAspAsn 26  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||  
Db 51291 GGAGTTGAAAAGAAAAC 51308  
RESULT 20  
AAS94455  
ID AAS94455 standard; CDNA; 3711 BP.  
XX  
AC AAS94455;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #30259.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR P-PSDB; ABG30268.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity -  
PT  
PT Claim 1: SEQ ID No 30259; 103pp; English.  
PS  
XX The invention relates to isolated polynucleotide (I) and  
polypeptide (II) sequences. (I) is useful as hybridisation probes,  
polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
and gene mapping, and in recombinant production of (II). The  
polynucleotides are also used in diagnostics as expressed sequence tags  
for identifying expressed genes. (I) is useful in gene therapy techniques  
to restore normal activity of (II) or to treat disease states involving  
(II). (II) is useful for generating antibodies against it, detecting or  
quantitating a polypeptide in tissue, as molecular weight markers and as  
a food supplement. (II) and its binding partners are useful in medical  
imaging of sites expressing (II). (I) and (II) are useful for treating  
disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 3711 BP; 848 A; 945 C; 1056 G; 862 T; 0 other;  
Alignment Scores:  
Pred. No.: 114 Length: 3711  
Score: 50.00 Matches: 11  
Percent Similarity: 66.67% Conservative: 5  
Best Local Similarity: 45.83% Mismatches: 8  
Query Match: 38.76% Indels: 0





Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x ABL25211 (1-16135)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
|||||  
Db 5971 CTGGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAAAGCGCGCAGTTGCTCTCTGTGGCA 5912  
|||||

QY 24 AlaAspAsn 26  
|||||

Db 5911 GCGGACAA 5903  
|||||

#### RESULT 15

ABL25210  
ID ABL25210 standard; DNA: 18603 BP.

XX AC ABL25210;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 27103.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX PS Claim 1; SEQ ID NO 27103; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 18603 BP; 4763 A; 4409 C; 4291 G; 5140 T; 0 other;

#### Alignment Scores:

Pred. No.: 388 Length: 18603  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x ABL25210 (1-186603)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
|||||  
Db 11563 CTGGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAAAGCGCGCAGTTGCTCTCTGTGGCA 11622  
|||||

QY 24 AlaAspAsn 26  
|||||

Db 11623 GCGGACAA 11631  
|||||

#### RESULT 16

ABL12683/C  
ID ABL12683 standard; cDNA: 1960 BP.

XX AC ABL12683;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 32531.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR P-PSDB; ABB68580.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX PS Claim 1; SEQ ID NO 32531; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1960 BP; 505 A; 526 C; 485 G; 444 T; 0 other;

#### Alignment Scores:

Pred. No.: 34 Length: 1960  
Score: 51.00 Matches: 11  
Percent Similarity: 53.85% Conservative: 3  
Best Local Similarity: 42.31% Mismatches: 12  
Query Match: 39.53% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x ABL12683 (1-19660)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21

PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154778.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157755.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Alignment Scores:  
Pred. No.: 12.7 Length: 1230  
Score: 52.00 Matches: 11  
Percent Similarity: 72.73% Conservative: 5  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 40.31% Indels: 0  
DB: 21 Gaps: 0  
  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAC33266 (1-1230)  
  
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValVallys 23  
||| |||||:::||||: ||| ||| :::::|||||::: |||  
Db 518 CTTCGGCAGCTCTCCGACAAAAGGACATGCTCAAAATTCAAAAGCTCAATGCAAG 459  
  
Qy 24 Alaasp 25  
|||||  
Db 458 CCGGAT 453  
  
RESULT 14  
ABL25211/C  
ID ABL25211 standard; DNA; 16135 BP.  
XX AC ABL25211;  
XX XX  
DT 26-MAR-2002 (first entry)  
XX XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 27106.  
XX XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX XX  
OS Drosophila melanogaster.  
XX XX  
FN WO200171042-A2.  
XX XX  
PD 27-SEP-2001.  
XX XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX XX  
PA (PEKE ) PE CORP NY.  
XX XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX XX  
DR WPI; 2001-656860/75.  
XX XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -  
XX XX  
PS Claim 1; SEQ ID NO 27106; 2lpp + Sequence Listing; English.  
XX XX  
CC The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.  
XX XX  
SQ Sequence 16135 BP; 4290 A; 3936 C; 4034 G; 3875 T; 0 other;  
  
Alignment Scores: 324 Length: 16135  
Pred. No.: 324

Alignment Scores: 9.78 Length: 1002  
Pred. No.: 52.00 Matches: 11  
Score: 72.73% Conservative: 5  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 40.31% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AAC54228 (1-1002)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
||| |||||:||||: ||| ||| :|||:||||: |||  
Db 442 CTTCCGCACTCTCCGACAAAGGACATCCCTCAATTCAAAAGCTGCATCGCAAG 383

QY 24 AlaAsp 25

|||||  
Db 382 GCGGAT 377

RESULT 13

AAC33266/c

ID AAC33266 standard; DNA; 1230 BP.

XX AC AAC33266;

XX DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 2396.  
XX KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;

XX KW metabolic pathway; promoter; termination sequence; ss.  
XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0123788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.

PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139434.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

```
XX
SQ Sequence 3349 BP; 808 A; 953 C; 851 G; 737 T; 0 other;

Alignment Scores:
Pred. No.: 20.1 Length: 3349
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservatives: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL13527 (1-3349)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
DB 1122 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGAGGATCGCGGCGTGGCGCC 1063
QY 25 AspAsnAlaAla 28
DB 1062 GATGCTGCGGCG 1051

RESULT 11
ABL13526/C
ID ABL13526 standard; cDNA; 6398 BP.
XX
AC ABL13526;
XX
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 35060.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI: 2001-656860/75.
DR P-PSDB; ABB69423.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 35060; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6398 BP; 1688 A; 1632 C; 1352 G; 1726 T; 0 other;

Alignment Scores:
```

```
Pred. No.: 45.3 Length: 6398
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservatives: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL13526 (1-6398)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
DB 2298 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGAGGATCGCGGCGTGGCGCC 2239
QY 25 AspAsnAlaAla 28
DB 2238 GATGCTGCGGCG 2227

RESULT 12
AAC54228/C
ID AAC54228 standard; DNA; 1002 BP.
XX
AC AAC54228;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 77144.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
```

CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.

XX  
SQ Sequence 666 BP; 250 A; 115 C; 119 G; 182 T; 0 other;

## Alignment Scores:

Pred. No.: 0.787 Length: 666  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservatve: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 18 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAT67891 (1-666)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
DB 229 ATACAGCCCTACAGGAGCAATTTGACGCTTTAGATTCTCAAGAAAAAGTCGTAGCAAA 288

QY 24 AlaAspAsn 26

DB 289 TGGGATAAC 297

## RESULT 9

AAX30461

ID AAX30461 standard; DNA; 1239 BP.

XX AC AAX30461;

XX DT 08-JUN-1999 (first entry)

XX H. pylori secreted protein ORF 09cel0413\_35336707\_f2\_9.  
XX Vaccine; probe; diagnostic; ORF; cell envelope protein;  
XX secreted protein; cellular protein; ds.

XX Helicobacter pylori.

XX WO9818323-A1.

XX PD 07-MAY-1998.

XX PF 28-OCT-1997; 97WO-US19575.

XX PR 14-JUL-1997; 97US-0891928.

XX PR 28-OCT-1996; 96US-0739150.

XX PR 06-DEC-1996; 96US-0759739.

XX (ASTR ) ASTRA AB.

XX PI Alm RA, Smith D;

XX DR WPI; 1998-271811/24.

XX DR P-PSDB; AAY10994.

XX Helicobacter pylori nucleic acids and proteins - used to develop  
PT products for the detection, prevention and treatment of H. pylori  
PT infections  
XX  
PS Claims 3, 4; Page 126-127; 279pp; English.  
XX Recombinant or substantially pure preparations of H. pylori polypeptides  
CC are disclosed, together with the nucleic acids encoding them. In all,  
CC 73 ORFs are shown. The proteins are variously cell envelope proteins,  
CC secreted proteins or other cellular proteins. Vaccines containing at least  
CC nucleic acids or proteins are claimed, as are probes containing at least  
CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful  
CC for treating or reducing the risk of H. pylori infections, and the

CC probes can be used diagnostically for detecting the presence of  
CC Helicobacter in a sample. The products are also of use in screening  
CC for compounds having the ability to interfere with the H. pylori life  
CC cycle or to inhibit H. pylori infection.

XX  
SQ Sequence 1239 BP; 472 A; 234 C; 206 G; 327 T; 0 other;

## Alignment Scores:

Pred. No.: 1.72 Length: 1239  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservatve: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 19 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAX30461 (1-1239)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
DB 145 ATACAGCCCTACAGGAGCAATTTGACGCTTTAGATTCTCAAGAAAAAGTCGTAGCAAA 204

QY 24 AlaAspAsn 26

DB 205 TGGGATAAC 213

## RESULT 10

ABL13527/C

ID ABL13527 standard; cDNA; 3349 BP.

XX AC ABL13527;

XX DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 35063.  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX DR P-PSDB; ABB69424.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 35063; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspClnThrAspAlaLeuGlnSerGluGluAlaAla 20  
Db 556 AGTGACGCTTAGAAGCATTCGGCGATCAACAGACGCTTTACATCAGAGAAGCTGCG 615  
Qy 21 ValValLysAlaAspAsnAlaAla 28  
Db 616 GTTGTTAAAGCGGATACGCTGCT 639  
RESULT 7  
AAT67618  
ID AAT67618 standard; DNA: 576 BP.  
XX AC AAT67618;  
XX 10-JUL-1997 (first entry)  
XX H. pylori secreted or periplasmic protein ORF 35336707.aa.  
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacterium; life cycle; activator;  
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
KW diagnosis; ds.  
XX Helicobacter pylori.  
XX Key Location/Qualifiers  
FH 1..576  
FT /\*tag= a  
FT /trans\_except= (pos: 541..543, aa: Glu)  
FT /trans\_except= (pos: 547..549, aa: Asp)  
FT /note= "no stop codon given"  
FT 559..561  
FT /\*tag= b  
FT /note= "encodes Asn"  
FT 574..576  
FT /\*tag= c  
FT /note= "encodes Asn"  
XX W09640893-A1.  
XX 19-DEC-1996.  
XX 06-JUN-1996; 96WO-US09122.  
XX 01-APR-1996; 96US-0630405.  
XX 07-JUN-1995; 95US-0487032.  
XX (ASTR ) ASTRA AB.  
XX Berglindh OT, Smith D, Mellgaard BL;  
XX WPI; 1997-052306/05.  
XX P-PSDB; AAW20445.  
XX Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
XX Claim 23; Page 268; 1481pp; English.  
XX This sequence encodes a H. pylori secreted or periplasmic protein.  
XX The protein may be used in a vaccine to prevent or treat H. pylori  
XX infection or to identify H. pylori polypeptide binding compounds,  
XX useful as potential H. pylori life cycle activators or inhibitors.  
XX The genomic sequence of H. pylori (ATCC 55679) was determined from  
XX overlapping contigs generated by mechanically shearing the bacterial  
XX DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
XX and the predicted coding regions defined by computer evaluation. To  
XX identify likely H. pylori antigens for vaccine development, the amino  
XX acid sequences predicted from various ORF were analysed for significant  
XX homology to other known or exported membrane proteins. Having identified  
XX and determined the sequences of interest, particular regions can be  
XX isolated from H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts.  
XX Sequence 576 BP; 217 A; 97 C; 95 G; 161 T; 6 other;  
Alignment Scores:  
Pred. No.: 0.655 Length: 576  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 18 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAT67618 (1-576)  
Qy 4 LeuGluAlaLeuAlaAspClnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
Db 145 ATACAGGCCCTACAGGACCAATTCAGCGCTTTAGATTCTCAAGAAAAACTCGTTAGCAAA 204  
Qy 24 AlaAspAsn 26  
Db 205 TCGGATAAC 213  
RESULT 8  
AAT67891  
ID AAT67891 standard; DNA: 666 BP.  
XX AC AAT67891;  
XX 14-JUL-1997 (first entry)  
XX H. pylori secreted or periplasmic protein ORF 02cel0216orf1.  
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacterium; life cycle; activator;  
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
KW diagnosis; ds.  
XX Helicobacter pylori.  
XX Key Location/Qualifiers  
FH 1..666  
FT /\*tag= a  
FT /note= "no stop codon given"  
XX W09640893-A1.  
XX 19-DEC-1996.  
XX 06-JUN-1996; 96WO-US09122.  
XX 01-APR-1996; 96US-0630405.  
XX 07-JUN-1995; 95US-0487032.  
XX (ASTR ) ASTRA AB.  
XX Berglindh OT, Smith D, Mellgaard BL;  
XX WPI; 1997-052306/05.  
XX P-PSDB; AAW20638.  
XX Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
XX Claim 23; Page 770-771; 1481pp; English.  
XX The present sequence encodes a H. pylori secreted or periplasmic protein.  
XX The protein may be used in a vaccine to prevent or treat H. pylori  
XX infection or to identify H. pylori polypeptide binding compounds,  
XX useful as potential H. pylori life cycle activators or inhibitors.  
XX The genomic sequence of H. pylori (ATCC 55679) was determined from  
XX overlapping contigs generated by mechanically shearing the bacterial  
XX DNA. The sequences were analysed for ORF of at least 180 nucleotides,



Qy 21 ValValLysAlaAspAsnAlaAa 28  
|||||  
Db 393 GTTGTTAAAGCGGATAACGCTGCT 416

RESULT 5  
AAD00561  
ID AAD00561 standard; DNA; 777 BP.  
AC AAD00561;  
XX  
XX  
XX  
XX 29-AUG-2000 (first entry)  
XX

XX Streptococcus pyogenes strain K7L9 partial GRAB protein encoding DNA.  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection; ds.  
XX Streptococcus pyogenes.  
XX

XX Key Location/Qualifiers  
FH 1..777  
FT CDS  
FT /\*tag= a  
FT /product= "GRAB protein"  
FT /note= "Does not include stop codon"  
FT /partial  
XX

PN WO200026240-A2.  
XX

XX 11-MAY-2000.  
XX

XX 02-NOV-1999; 99WO-GB03631.  
XX

XX 02-NOV-1998; 98GB-0023975.  
XX

XX (ACTI-) ACTINOVA LTD.  
XX

XX Björck LH, Rasmussen M;  
XX

XX WPI; 2000-365572/31.  
DR

XX P-PSDB; AAY71043.  
DR

XX New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX

XX Claim 13; Page 63-64; 67pp; English.  
XX

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a DNA encoding partial GRAB protein from S. pyogenes strain K7L9.  
CC The protein has alpha2M binding region and is useful in vaccine  
XX composition.  
XX

SQ Sequence 777 BP; 269 A; 151 C; 178 G; 179 T; 0 other;

Alignment Scores:  
Pred. No.: 2,74e-13 Length: 777  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAD00561 (1-777)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAa 20

Db 319 AGTGACGCCCTTAGAAGCATTTGGCGGATCAACACAGCGCTTTACAATCAGAGAAGCTGCG 378  
|||||

Qy 21 ValValLysAlaAspAsnAlaAa 28  
|||||

Db 379 GTTGTTAAAGCGGATAACGCTGCT 402  
|||||

RESULT 6  
AAD00563

ID AAD00563 standard; DNA; 853 BP.  
XX

XX AAD00563;  
XX

XX 29-AUG-2000 (first entry)  
XX

XX Streptococcus pyogenes strain AP49 partial GRAB protein encoding DNA.  
XX

XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection; ds.  
XX

XX Streptococcus pyogenes.  
XX

XX Key Location/Qualifiers  
FH 37..852  
FT CDS  
FT /\*tag= a  
FT /product= "GRAB protein"  
FT /partial  
XX

PN WO200026240-A2.  
XX

XX 11-MAY-2000.  
XX

XX 02-NOV-1999; 99WO-GB03631.  
XX

XX 02-NOV-1998; 98GB-0023975.  
XX

XX (ACTI-) ACTINOVA LTD.  
XX

XX Björck LH, Rasmussen M;  
XX

XX WPI; 2000-365572/31.  
DR

XX P-PSDB; AAY71045.  
DR

XX New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
XX Claim 13; Page 64; 67pp; English.  
XX

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a DNA encoding partial GRAB protein from S. pyogenes strain AP49.  
CC The protein has alpha2M binding region and is useful in vaccine  
XX composition.  
XX

SQ Sequence 853 BP; 295 A; 171 C; 197 G; 190 T; 0 other;

Alignment Scores:  
Pred. No.: 3.08e-13 Length: 853  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAD00563 (1-853)

```

FH Key      Location/Qualifiers
FT CDS      1..654
FT          /*tag= a
FT          /product= "GRAB protein"
FT sig_peptide
FT          /*tag= b
FT mat_peptide
FT          /*tag= c
FT          /product= "Mature GRAB protein"
XX
PN WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACTI-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
XX
XX P-PSDB; AAY71042.
XX
XX New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein -
XX
XX Claim 13; Page 63; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G Streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A Streptococcus. Antibodies against GRAB
XX are useful for treating Streptococcus pyogenes infections. The protein
XX is also useful for purifying alpha2M from a sample. The present sequence
XX is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.
XX
XX Sequence 654 BP; 234 A; 118 C; 143 G; 159 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 2 21e-13 Length: 654
XX Score: 129.00 Matches: 28
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-847-539A-6_COPY_59_86 (1-28) x AAD00560 (1-654)
XX
XX Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20
XX 274 TCAGATGCGCTTAGAGCATTAGCGGATCAACAGACGCTTTACATCAGAGAGCTGCG 333
XX
XX Qy 21 ValValLysAlaAspAsnAlaLa 28
XX 334 GTTGTAAAGCGGATAGCGTGCT 357
XX
XX RESULT 4
XX AAD00559
XX ID AAD00559 standard; DNA; 764 BP.
XX
XX AC AAD00559;
XX
XX 29-AUG-2000 (first entry)
XX
XX Streptococcus pyogenes strain SF370 GRAB protein encoding DNA.
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX immune response; Streptococcus pyogenes infection; ds.

```

```

XX Streptococcus pyogenes.
XX
XX Key      Location/Qualifiers
XX CDS      60..713
XX          /*tag= a
XX          /product= "GRAB protein"
XX sig_peptide
XX          /*tag= b
XX mat_peptide
XX          /*tag= c
XX          /product= "Mature GRAB protein"
XX primer_bind
XX          /*tag= d
XX          /bound_moiety= "Primer 1 represented in AAD00565"
XX primer_bind
XX          /*tag= e
XX          /bound_moiety= "Primer 2 represented in AAD00566"
XX primer_bind
XX          /*tag= f
XX          /bound_moiety= "Primer 3 represented in AAD00567"
XX primer_bind
XX          /*tag= g
XX          /bound_moiety= "Primer 4 represented in AAD00568"
XX primer_bind
XX          /*tag= h
XX          /bound_moiety= "Primer 5 represented in AAD00569"
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACTI-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
XX
XX P-PSDB; AAY71042.
XX
XX New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein -
XX
XX Example 1; Fig 2B; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G Streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A Streptococcus. Antibodies against GRAB
XX are useful for treating Streptococcus pyogenes infections. The protein
XX is also useful for purifying alpha2M from a sample. The present sequence
XX is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.
XX
XX Sequence 764 BP; 279 A; 131 C; 159 G; 195 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 2 68e-13 Length: 764
XX Score: 129.00 Matches: 28
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-847-539A-6_COPY_59_86 (1-28) x AAD00559 (1-764)
XX
XX Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20
XX 333 TCAGATGCGCTTAGAGCATTAGCGGATCAACAGACGCTTTACATCAGAGAGCTGCG 392
XX
XX Db

```

```

FT      /*tag= a
FT      /product= "GRAB protein"
FT      /partial
XX
XX      WO200026240-A2.
XX
XX      11-MAY-2000.
XX
XX      02-NOV-1999; 99WO-GB03631.
XX
XX      02-NOV-1998; 98GB-0023975.
XX
XX      (ACTI-) ACTINOVA LTD.
XX
XX      Bjorck LH, Rasmussen M;
XX
XX      WPI; 2000-365572/31.
XX
XX      P-PSDB; AAY71044.
XX
XX      New alpha2M binding protein for generating a protective immune response
XX      to group A streptococcus and purifying the binding protein
XX
XX      Claim 13; Page 64; 67pp; English.
XX
XX      The patent discloses a new family of proteins termed GRAB (protein G
XX      related alpha2M binding protein) from Streptococcus pyogenes which have
XX      the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX      protein G of group G Streptococcus. GRAB protein and peptides derived
XX      from it are used in vaccine compositions for generating a protective
XX      immune response against group A Streptococcus. Antibodies against GRAB
XX      are useful for treating Streptococcus pyogenes infections. The protein
XX      is also useful for purifying alpha2M from a sample. The present sequence
XX      is a DNA encoding partial GRAB protein from S. pyogenes strain AP1.
XX      The protein has alpha2M binding region and is useful in vaccine
XX      composition.
XX
XX      Sequence 469 BP; 180 A; 90 C; 99 G; 100 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.: 1.45e-13 Length: 469
XX      Score: 129.00 Matches: 28
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 100.00% Indels: 0
XX      DB: 21 Gaps: 0
XX
XX      US-09-847-539A-6_COPY_59_86 (1-28) x AAD00562 (1-469)
XX
XX      Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 20
XX      Db 172 TCAGATGCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAATCAGAGAAGCTGCG 231
XX
XX      Qy 21 ValValLysAlaAspAsnAlaAla 28
XX      Db 232 GTTGTAAAGCGGATACGCTGCT 255
XX
XX      RESULT 2
XX      AAD00564
XX      ID AAD00564 standard; DNA; 504 BP.
XX
XX      AC AAD00564;
XX
XX      Dt 29-AUG-2000 (first entry)
XX
XX      Streptococcus pyogenes strain KTL3 partial GRAB protein encoding DNA.
XX
XX      GRAB protein; protein G related alpha2M binding protein; vaccine;
XX      alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX      immune response; Streptococcus pyogenes infection; ds.
XX
XX      Streptococcus pyogenes.
XX
XX      Key Location/Qualifiers

```

```

CDS      1..504
FT      /*tag= a
FT      /product= "GRAB protein"
FT      /partial
XX
XX      WO200026240-A2.
XX
XX      11-MAY-2000.
XX
XX      02-NOV-1999; 99WO-GB03631.
XX
XX      02-NOV-1998; 98GB-0023975.
XX
XX      (ACTI-) ACTINOVA LTD.
XX
XX      Bjorck LH, Rasmussen M;
XX
XX      WPI; 2000-365572/31.
XX
XX      P-PSDB; AAY71046.
XX
XX      New alpha2M binding protein for generating a protective immune response
XX      to group A streptococcus and purifying the binding protein
XX
XX      Claim 13; Page 65; 67pp; English.
XX
XX      The patent discloses a new family of proteins termed GRAB (protein G
XX      related alpha2M binding protein) from Streptococcus pyogenes which have
XX      the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX      protein G of group G Streptococcus. GRAB protein and peptides derived
XX      from it are used in vaccine compositions for generating a protective
XX      immune response against group A Streptococcus. Antibodies against GRAB
XX      are useful for treating Streptococcus pyogenes infections. The protein
XX      is also useful for purifying alpha2M from a sample. The present sequence
XX      is a DNA encoding partial GRAB protein from S. pyogenes strain KTL3.
XX      The protein has alpha2M binding region and is useful in vaccine
XX      composition.
XX
XX      Sequence 504 BP; 188 A; 97 C; 108 G; 111 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.: 1.59e-13 Length: 504
XX      Score: 129.00 Matches: 28
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 100.00% Indels: 0
XX      DB: 21 Gaps: 0
XX
XX      US-09-847-539A-6_COPY_59_86 (1-28) x AAD00564 (1-504)
XX
XX      Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 20
XX      Db 208 TCAGATGCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAATCAGAGAAGCTGCG 267
XX
XX      Qy 21 ValValLysAlaAspAsnAlaAla 28
XX      Db 268 GTTGTAAAGCGGATACGCTGCT 291
XX
XX      RESULT 3
XX      AAD00560
XX      ID AAD00560 standard; DNA; 654 BP.
XX
XX      AC AAD00560;
XX
XX      Dt 29-AUG-2000 (first entry)
XX
XX      Streptococcus pyogenes strain SF370 GRAB protein coding region.
XX
XX      GRAB protein; protein G related alpha2M binding protein; vaccine;
XX      alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX      immune response; Streptococcus pyogenes infection; ds.
XX
XX      Streptococcus pyogenes.
XX
XX      Key Location/Qualifiers

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:10:56 : Search time 40.4278 Seconds  
(without alignments)  
1189.122 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86

Perfect score: 129  
Sequence: 1 SDALPALADQTDALQSEAVVVKADNAA 28

Scoring table: BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_092548\_3413/app\_query.fasta\_1.526  
-DB=N\_Geneseq\_032802 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=40  
-MODE=LOCAL -OUTFMT=ptc -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539\_032802\_1.48 -runat\_10102002\_092548\_3413 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAP -LARGESIZE -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129	100.0	469	21	Streptococcus pyog
2	129	100.0	504	21	Streptococcus pyog
3	129	100.0	554	21	Streptococcus pyog
4	129	100.0	764	21	Streptococcus pyog
5	129	100.0	777	21	Streptococcus pyog
6	129	100.0	853	21	Streptococcus pyog
7	57	44.2	576	18	H. pylori secreted
8	57	44.2	566	18	H. pylori secreted
9	57	44.2	1239	19	H. pylori secreted
C 10	54	41.9	3349	23	Drosophila melanog
C 11	54	41.9	6398	23	Drosophila melanog
C 12	52	40.3	1002	21	Arabidopsis thalia
C 13	52	40.3	1230	21	Arabidopsis thalia
C 14	52	40.3	16135	23	Drosophila melanog
C 15	52	40.3	18603	23	Drosophila melanog
C 16	51	39.5	1960	23	Drosophila melanog
C 17	51	39.5	2618	23	Drosophila melanog
C 18	51	39.5	4221	23	Drosophila melanog
19	51	39.5	2944528	24	Listeria monocytog
20	50	38.8	3711	23	DNA encoding novel
C 21	48	37.2	1616	23	Drosophila melanog
C 22	48	37.2	4229	23	Drosophila melanog
C 23	47	36.4	4403765	22	Mycobacterium tube
C 24	47	36.4	4411529	22	Mycobacterium tube
25	46	35.7	636	21	Aspergillus niger
26	46	35.7	2115	22	C glutamicum codin
C 27	46	35.7	2160	22	C glutamicum codin
C 28	46	35.7	69936	21	N. meningitidis pa
C 29	46	35.7	349980	21	Neisseria meningit
30	46	35.7	349980	21	C glutamicum codin
31	46	35.7	1437668	21	N. meningitidis B
32	45	34.9	417	21	Escherichia coli f
C 33	45	34.9	461	21	Fusarium venenatum
C 34	45	34.9	1365	21	Escherichia coli f
C 35	45	34.9	1368	21	Escherichia coli f
36	45	34.9	1380	21	Escherichia coli f
37	45	34.9	1380	21	Escherichia coli f
38	45	34.9	1383	21	Escherichia coli f
39	45	34.9	1383	21	Escherichia coli f
C 40	45	34.9	1773	21	Coding sequence of
C 41	45	34.9	2145	21	Escherichia coli f
C 42	45	34.9	2599	21	Escherichia coli f
C 43	45	34.9	4521	20	Murine pcip gene (
C 44	45	34.9	4860	20	Murine p/cip gene,
45	45	34.9	5091	23	Drosophila melanog

ALIGNMENTS

RESULT 1  
AAD00562  
ID AAD00562 standard; DNA; 469 BP.  
XX  
AC AAD00562;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain APl partial GRAB protein encoding DNA.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW Immune response; Streptococcus pyogenes infection; ds.  
XX  
OS Streptococcus pyogenes.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..468

THIS PAGE BLANK (uspto)

J. Bacteriol. 173, 7650-7664, 1991  
A:Title: Molecular and genetic analysis of a region of plasmid pCF10 containing positive  
terococcus faecalis.

A:Reference number: A41662; MUID:92041679

A:Accession: G41662

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <KAO>

A:Cross-references: GB:M64978; NID:g150552; PIDN:AAA65847.1; PID:g150554

C:Genetics:

A:Genome: plasmid

Query Match

Best Local Similarity 9.4%; Score 73; DB 2; Length 891;

Matches 21; Conservative 20; Mismatches 58; Indels 0; Gaps 0;

Qy 33 ERAIDELKKQAIEDKEATTAIEAASSDALEALADQTDALQSEAAVVKADNAASDALEA 92

Db 66 KQAVDQKQVADVAKKEKDAIDQSVKQDQAVVQDKNDALDQSQQAVTDQAVVDEAKKV 125

Qy 93 LADQTDALQSEAEVQSDNAASDAWEKAATPIALDVKK 131

Db 126 VDEATPSAIERAKQGVATDQAVDEQKQVVDQAOQTDVNG 164

RESULT 39

T14968

phage lambda-related protein - Yersinia pestis plasmid pMT1

C:Species: Yersinia pestis

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 22-Oct-1999

C:Accession: T14968

R:Lindler, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.

Infect. Immun. 66, 5731-5742, 1998

A:Title: Complete DNA sequence and detailed analysis of the Yersinia pestis KIM5 plasmid

A:Reference number: Z18268; MUID:99043898

A:Accession: T14968

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-978 <LIN>

A:Cross-references: EMBL:AF074611; NID:g3883003; PID:g3883051; PIDN:AAC82711.1

C:Genetics:

A:Gene: Y1051

A:Genome: plasmid pMT1

Query Match

Best Local Similarity 9.4%; Score 73; DB 2; Length 978;

Matches 20; Conservative 12; Mismatches 28; Indels 0; Gaps 0;

Qy 56 AASSDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAEVQSDNAAS 115

Db 237 AAKTSEDNSAASETNAESKAAALSSANSASEALQYAESAKTSKEAAAAAEEAAAAAN 296

RESULT 40

A54817

ATPase ScfII, chromosomal scaffold - chicken

C:Species: Gallus gallus (chicken)

C:Date: 12-Apr-1995 #sequence\_revision 12-Apr-1995 #text\_change 02-Feb-2001

C:Accession: A54817

R:Saitoh, N.; Goldberg, I.G.; Wood, E.R.; Earnshaw, W.C.

J. Cell Biol. 127, 303-318, 1994

A:Title: ScfII: an abundant chromosome scaffold protein is a member of a family of putative

A:Reference number: A54817; MUID:95014725

A:Accession: A54817

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1189 <SAI>

A:Cross-references: GB:X80792; NID:g572691; PIDN:CAA56767.1; PID:g572692

C:Superfamily: chromosome segregation protein SMC1

C:Keywords: nucleotide binding; P-loop

F:32-39/Region: nucleotide-binding motif A (P-loop)

Query Match

9.4%; Score 73; DB 2; Length 1189;

Best Local Similarity 24.7%; Pred. No. 12;  
Matches 21; Conservative 18; Mismatches 46; Indels 0; Gaps 0;  
Qy 62 LEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAEVQSDNAASDAWEKA 121  
Db 819 VEALVLELEQLKQEQASYKQOSEAAQQAIAISLKQVSALEAEAVKTRSLKNAENELSS 878  
Qy 122 ATPIALDVKKTKDTKPVVKKEERQN 146  
Db 879 KGLMEERTKDIKAKSAKIEKYEQN 903

Search completed: October 13, 2002, 02:12:00

Job time : 46.5134 secs

```
QY 26 EKIALRNEERAIDELKKQAIEDKEATTATIAEASSDALEALADQTDALQSEEAHVVKADNA 85
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 27 ENVVAKENNTSEGEKQNTVAETITTSVEAKETFPVEPTKTPAVQPEVAVVESSSA 86
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 86 ASDALEMALADQTDALQSEEAHVVDSDNAASDAWEKAATPIALDVK 130
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 87 DAGEAAVVAPEKVENAATNAEAKVEAVAAPKEVAVAEKK 131
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 34
C90769
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: C90769
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.-G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C90769
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-439 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA834546.1; PID:g13360583; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs1123

Query Match 9.4%; Score 73; DB 2; Length 439;
Best Local Similarity 35.3%; Pred. No. 4.1;
Matches 24; Conservative 7; Mismatches 37; Indels 0; Gaps 0;

QY 50 ATTAIEASSDALEALADQTDALQSEEAHVVKADNAASDALEALADQTDALQSEEAHVQ 109
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 150 ADTSAGDASERQAESAASAAKQSEERSSSSASAAKAKASESSQSAEAELSRRTAES 209
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 110 SDNAASDA 117
      : : : : :
Db 210 AGNAARDA 217
      : : : : :

RESULT 35
D84325
Htr17 transducer [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: D84325
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabil
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: D84325
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-536 <STO>
A:Cross-references: GB:AE004437; NID:g10581193; PIDN:AAG19968.1; GSPDB:GN00138
C:Genetics:
A:Gene: htr17
C:Superfamily: Halobacterium salinarum transducer protein htrI

Query Match 9.4%; Score 73; DB 2; Length 536;
Best Local Similarity 27.0%; Pred. No. 5;
Matches 24; Conservative 12; Mismatches 53; Indels 0; Gaps 0;

QY 39 ELKKQAIEDKEATTATIAEASSDALEALADQTDALQSEEAHVVKADNAASDALEALADQTD 98
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 EVKELAEETNEATDEITFLVDNLLSSQSVVDDSEQMAATVSSCTETVENALTAELEIGD 427
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 99 ALQSEEAHVVDSDNAASDAWEKAATPIAL 127
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 HVTVDNEAVQIIDDAQQAASADTTVSM 456
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 36
T33170
hypothetical protein M01B12.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33170
R:Becker, M.; Graves, T.; Ozersky, P.
submitted to the EMBL Data Library, May 1998
A:Description: The sequence of C. elegans cosmid M01B12.
A:Reference number: Z21297
A:Accession: T33170
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-719 <BEC>
A:Cross-references: EMBL:AF067624; PIDN:AAC17562.1; GSPDB:GN00019; CESP:M01B12.4
A:Experimental source: strain Bristol N2; clone M01B12
C:Genetics:
A:Gene: CESP:M01B12.4
A:Map position: 1
A:Introns: 5/3; 41/2; 135/3; 234/1; 438/3; 590/2; 626/1

Query Match 9.4%; Score 73; DB 2; Length 719;
Best Local Similarity 25.8%; Pred. No. 6.9;
Matches 24; Conservative 12; Mismatches 57; Indels 0; Gaps 0;

QY 25 PEKIALRNEERAIDELKKQAIEDKEATTATIAEASSDALEALADQTDALQSEEAHVVKADN 84
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 PRKLPECKPEARKSAPVAAQETAPVVVVPVVEAIPEVPAQVEEVLTVPEEPAAREA 313
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 85 AASDALEALADQTDALQSEEAHVVDSDNAASDA 117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 VVEEAPVAVVEEITVPAEPAVVAEEQQIAPEA 346
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 37
T14650
tail fiber protein homolog - Yersinia pestis plasmid pMT1
C:Species: Yersinia pestis
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14650
R:Hu, P.; Elliott, J.; McCready, P.; Skowronski, E.; Garnes, J.; Kobayashi, A.; Carra
submitted to the EMBL Data Library, March 1998
A:Description: Structural organization of virulence determinants in three Yersinia pe
A:Reference number: Z18168
A:Accession: T14650
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-860 <HUP>
A:Cross-references: EMBL:AF053947; NID:g2996286; PID:g2996320; PIDN:AAC13200.1
C:Genetics:
A:Gene: plasmid pMT1

Query Match 9.4%; Score 73; DB 2; Length 860;
Best Local Similarity 33.3%; Pred. No. 8.4;
Matches 20; Conservative 12; Mismatches 28; Indels 0; Gaps 0;

QY 56 AASDALEALADQTDALQSEEAHVVKADNAASDALEALADQTDALQSEEAHVVDSDNAAS 115
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 AAKTSEDNSAASETNAAESKAAALSSSSANSASALQVAESA KTSKEAAAAAEEAAAAN 178
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 38
G41662
130K surface exclusion protein Sec10 precursor - Enterococcus faecalis plasmid pCF10
C:Species: Enterococcus faecalis
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Oct-1999
C:Accession: G41662
R:Kao, S.M.; Olmsted, S.B.; Viksnins, A.S.; Gallo, J.C.; Dunny, G.M.
```







Db 210 ACNAARDA 217

RESULT 23

A33430

N:Caldesmon - chicken

C:Species: Gallus gallus (chicken)

C:Date: 27-Feb-1990 #sequence\_revision 27-Jun-1994 #text\_change 22-Jun-1999

C:Accession: A33430; A32642; A32445; A41064; A60461; PC2003; PX0022

R:Hayashi, K.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.

Biochem. Biophys. Res. Commun. 164, 503-511, 1989

A:Title: Primary structure and functional expression of h-caldesmon complementary DNA.

A:Reference number: A33430; MUID:90026426

A:Accession: A33430

A:Molecule type: mRNA

A:Residues: 1-771 <HAY>

A:Cross-references: GB:M28417; NID:g211895; PIDN:AAA48810.1; PID:g211896

A:Experimental source: gizzard

A:Note: Part of this sequence was confirmed by protein sequencing

R:Bryan, J.; Inai, M.; Lee, R.; Moore, P.; Cook, R.G.; Lin, W.G.

J. Biol. Chem. 264, 13873-13879, 1989

A:Title: Cloning and expression of a smooth muscle caldesmon.

A:Reference number: A32642; MUID:89340480

A:Accession: A32642

A:Molecule type: mRNA

A:Residues: 1-318,334-771 <BRY>

A:Cross-references: GB:J04968; NID:g212656; PIDN:AAA49067.1; PID:g212657

A:Note: The authors translated the codon GAA for residue 743 as Leu

R:Hayashi, K.; Yamada, S.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.

Biochem. Biophys. Res. Commun. 161, 38-45, 1989

A:Title: 35kDa fragment of h-caldesmon conserves two consensus sequences of the tropomyosin

A:Reference number: A32445; MUID:89273666

A:Accession: A32445

A:Molecule type: mRNA

A:Residues: 466-771 <HA2>

A:Cross-references: GB:M26684; NID:g211897; PIDN:AAA48811.1; PID:g211898

R:Wang, C.L.A.; Wang, L.W.C.; Lu, R.C.

Biochem. Biophys. Res. Commun. 162, 746-752, 1989

A:Title: Caldesmon has two calmodulin-binding domains.

A:Reference number: A60461; MUID:89334885

A:Accession: A60461

A:Molecule type: protein

A:Residues: 597-600;678-696;711-721 <HAK>

R:Wang, C.L.A.; Wang, L.W.C.; Lu, R.C.

Biochem. Biophys. Res. Commun. 162, 746-752, 1989

A:Title: Caldesmon has two calmodulin-binding domains.

A:Reference number: A60461; MUID:89334885

A:Accession: A60461

A:Molecule type: protein

A:Residues: 2-17, 'X', 19-38;466-485 <WAN>

R:Haruna, M.; Hayashi, K.; Yano, H.; Takeuchi, O.; Sobue, K.

Biochem. Biophys. Res. Commun. 197, 145-153, 1993

A:Title: Common structural and expressional properties of vertebrate caldesmon genes.

A:Reference number: PC2003; MUID:94071934

A:Accession: PC2003

A:Molecule type: DNA

A:Residues: 74-419 <HAR>

R:Takagi, T.; Yazawa, M.; Ueno, T.; Suzuki, S.; Yagi, K.

J. Biochem. 106, 778-783, 1989

A:Title: Amino acid sequence studies on cyanogen bromide peptides of chicken caldesmon

A:Reference number: PX0022; MUID:90130380

A:Accession: PX0022

A:Molecule type: protein

A:Residues: 462-477, 'D', 479-563;674-762, 'A', 763-771 <TAK>

C:Comment: This protein plays a vital role in the regulation of smooth muscle and nonmuscle

C:Comment: The binding of caldesmon to F-actin is modulated by calcium and calmodulin.

C:Comment: Two calmodulin molecules can bind to nonoverlapping domains of each caldesmon

C:Superfamily: caldesmon

C:Keywords: actin binding; alternative splicing; calmodulin binding; cytoskeleton; muscle

F:1-771/Product: h-caldesmon #status predicted <HAY>

F:1-318,334-771/Product: h-caldesmon, alternative splice form #status predicted <LMAT

F:266-390/Region: 13-residue repeats

F:511-582/Region: tropomyosin binding

F:622-636/Region: tropomyosin binding

F:597,682,717/Binding site: phosphate (Ser) (covalent) (by cdc2 kinase) #status exper

F:688,711/Binding site: phosphate (Thr) (covalent) (by cdc2 kinase) #status experimen

Query Match 10.0%; Score 78; DB 1; Length 771;

Best Local Similarity 26.4%; Pred. No. 2.1; Mismatches 17; Indels 0; Gaps 0;

Matches 24; Conservative 24; Mismatches 17; Indels 0; Gaps 0;

QY 26 EKALNREERAIDELKKQATIEDKATTAIEAASDSDALEALADQTDALQSEAAVVKADNA 85

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 296 EKRAAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKA 355

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 86 ASDALEALADQTDALQSEAAVVKASD 116

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 356 ABERAKAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKAE 386

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 24

IS1116

NF-180 - sea lamprey

C:Species: Petromyzon marinus (sea lamprey)

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000

C:Accession: IS1116

R:Jacobs, A.J.; Kamholz, J.; Selzer, M.E.

Brain Res. Mol. Brain Res. 29, 43-52, 1995

A:Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation

A:Reference number: IS1116; MUID:95287814

A:Accession: IS1116

A>Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: mRNA

A:Residues: 1-1110 <JAC>

A:Cross-references: EMBL:U19361; NID:g632548; PIDN:AAA80106.1; PID:g632549

C:Superfamily: neurofilament triplet H protein

Query Match 9.9%; Score 77; DB 2; Length 1110;

Best Local Similarity 23.1%; Pred. No. 4.1; Mismatches 24; Indels 0; Gaps 0;

Matches 21; Conservative 24; Mismatches 24; Indels 0; Gaps 0;

QY 55 EAASDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKSDNA 114

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 706 EAEEVEVKEEVTKSDAE 765

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 115 SDAAWEAKATPIALDVKKTKDTKTPVVKKEERQ 145

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 766 EBAEAEASDDEKPEEVEKESAPVAPEAKK 796

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 25

T22976

Hypothetical protein F59A2.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T22976; T23157

R:Lightning, J.

submitted to the EMBL Data Library, June 1994

A:Reference number: Z19645

A:Accession: T22976

A>Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-1133 <WIL>

A:Cross-references: EMBL:Z34801; PIDN:CAA84332.1; GSPDB:GN00021; CESP:F59A2.6

A:Experimental source: clone F59A2

R:Burton, J.

submitted to the EMBL Data Library, October 1995

A:Reference number: Z19700

A:Accession: T23157

A>Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-1133 <WIL>

A:Cross-references: EMBL:Z66514; PIDN:CAA91344.1; GSPDB:GN00021; CESP:F59A2.6

A:Molecule type: DNA  
A:Residues: 1-507 <FUE>

A:Cross-references: GB:X16421; EMBL:M26048; NID:g43333; PIDN:CAA34442.1; PID:g43334  
A:Note: the authors translated the codon CGT for residues 221 and 223 as Lys

Query Match 10.2%; Score 79; DB 2; Length 507;

Best Local Similarity 16.9%; Pred. No. 1;  
Matches 21; Conservative 30; Mismatches 73; Indels 0; Gaps 0;

Qy 26 EKALRNEERAIDELKKOAYEDKCATTAIRAASDDALEALADQTDALQSEAAVVKADNA 85

Db 218 EQAIRQEARLAERQARQAEKAEKAREQAQAQAQTALSSASTTSSSSAAQSSE 277

Qy 86 ASDALEALADQTDALQSEAAVWQSDNAASDAWEKAATPIALDVKKTKDKPVVKKERQ 145

Db 278 ESKAPESSTTEESPTSTSTTTSSSTGSSSTSSSTESSTVPSSTQESTPANTESSSS 337

Qy 146 NVNT 149

Db 338 SSNT 341

RESULT 19

S66805 hypothetical protein YOL109w - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein O0738

C:Species: *Saccharomyces cerevisiae*

C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 06-Feb-1998

C:Accession: S66805

R:Durant, P.; Hilger, F.; Portetelle, D.; Vandenbol, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66791

A:Molecule type: DNA

A:Residues: 1-113 <DUR>

A:Cross-references: EMBL:Z74851; NID:g1419974; PID:e251901; PID:g1419975; MIPS:YOL109w

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:Z601

A:Cross-references: SGD:S0005469; MIPS:YOL109w

A:Map position: 15L

Query Match 10.0%; Score 78; DB 2; Length 113;

Best Local Similarity 37.3%; Pred. No. 0.26;

Matches 19; Conservative 10; Mismatches 22; Indels 0; Gaps 0;

Qy 81 KADNAASDALEALADQTDALQSEAAVQSDNAASDAWEKAATPIALDVKK 131

Db 7 KAETAAODVOOKLEETKESLQKGOEVKEQAASIDNLKNEATPEAEQVK 57

RESULT 20

G90907

Probable tail fiber protein [Imported] - *Escherichia coli* (strain O157:H7, substrain RIM

C:Species: *Escherichia coli*

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: G90907

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: G90907

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-407 <HAY>

A:Cross-references: GB:BA000007; PIDN:BAE35654.1; PID:g13361697; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECS2231

Query Match

Best Local Similarity 10.0%; Score 78; DB 2; Length 407;

Matches 25; Conservative 36; Mismatches 36; Indels 0; Gaps 0;

Matches 25; Conservative 7; Mismatches 36; Indels 0; Gaps 0;

Qy 50

Db 148

Qy 110

Db 208

RESULT 21

E90968

probable tail fiber protein [Imported] - *Escherichia coli* (strain O157:H7, substrain

C:Species: *Escherichia coli*

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: E90968

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: E90968

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-437 <HAY>

A:Cross-references: GB:BA000007; PIDN:BAE36140.1; PID:g13362185; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECS2717

Query Match

Best Local Similarity 10.0%; Score 78; DB 2; Length 437;

Matches 25; Conservative 7; Mismatches 36; Indels 0; Gaps 0;

Qy 50

Db 148

Qy 110

Db 208

RESULT 22

E85816

probable tail fiber protein of prophage CP-933U Z3074 [Imported] - *Escherichia coli* (

C:Species: *Escherichia coli*

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: E85816

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206351

A:Accession: E85816

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-439 <STO>

A:Cross-references: GB:AE005174; NID:gl2516091; PIDN:AAE56993.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z3074

Query Match

Best Local Similarity 10.0%; Score 78; DB 2; Length 439;

Matches 25; Conservative 36; Mismatches 36; Indels 0; Gaps 0;

Qy 50

Db 150

Qy 110



[illegible]

Best Local Similarity 39.7%; Pred. No. 0.012;  
Matches 27; Conservative 8; Mismatches 33; Indels 0; Gaps 0;

QY 50 ATTATGAASSDALEALADOTDQLOSEAAVVKADNAASDALEALADOTDQLOSEAEVQ 109

Db 149 ADTSAEDASARQAESAASAKKSEASSSSSASEAAQAKASELSQATDAELSKKTAESA 208

QY 110 SDNAASDA 117

Db 209 AGNAARDA 216

#### RESULT 6

Probable tail component of prophage CP-933K Z0982 [imported] - Escherichia coli (strain

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: F85584

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: F85584

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-440 <STO>

A:Cross-references: GB:AE005174; NID:gl2513753; PIDN:AAQ55138.1; GSPDB:GN00145; UNGP:209

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z0982

Query Match 12.4%; Score 96; DB 2; Length 440;

Best Local Similarity 39.7%; Pred. No. 0.012;

Matches 27; Conservative 8; Mismatches 33; Indels 0; Gaps 0;

QY 50 ATTATGAASSDALEALADOTDQLOSEAAVVKADNAASDALEALADOTDQLOSEAEVQ 109

Db 151 ADTSAEDASARQAESAASAKKSEASSSSSASEAAQAKASELSQATDAELSKKTAESA 210

QY 110 SDNAASDA 117

Db 211 AGNAARDA 218

#### RESULT 7

S55890

plasma protein receptor MAG precursor - Streptococcus dysgalactiae

C:Species: Streptococcus dysgalactiae

C:Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 17-Mar-2000

C:Accession: S55890

R:Jonsson, H.; Frykberg, L.; Rantanen, L.; Guss, B.

Gene 143, 85-89, 1994

A:Title: MAG, a novel plasma protein receptor from Streptococcus dysgalactiae.

A:Reference number: S55890; MUID:94259307

A:Accession: S55890

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-413 <JON>

A:Cross-references: EMBL:L27798; NID:g475117; PIDN:AAA26921.1; PID:g475118

C:Superfamily: M5 protein

Query Match 11.7%; Score 91; DB 2; Length 413;

Best Local Similarity 59.4%; Pred. No. 0.04;

Matches 19; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 128 DYKTKDTPVVKKEERQNVNLTPTTGEESNP 159

Db 357 DPKKEDAKKPEAKKEAKAATLPTTGESNP 368

#### RESULT 8

H90854

Probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: H90854

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: H90854

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-437 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA835231.1; PID:gl3361273; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECS1808

Query Match 11.5%; Score 89; DB 2; Length 437;

Best Local Similarity 39.7%; Pred. No. 0.071;

Matches 27; Conservative 7; Mismatches 34; Indels 0; Gaps 0;

QY 50 ATTATGAASSDALEALADOTDQLOSEAAVVKADNAASDALEALADOTDQLOSEAEVQ 109

Db 148 ADTSAEDASARQAESAASAKKSEASSSSSASEAAQAKASELSQATDAELSKKTAESA 207

QY 110 SDNAASDA 117

Db 208 AGNAARDA 215

#### RESULT 9

EB8EQ3

Immediate-early protein - saimiriine herpesvirus 1 (strain 11)

C:Species: saimiriine herpesvirus 1

A:Note: host Saimiri sciureus (common squirrel monkey)

C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 16-Jul-1999

C:Accession: G36813; S20244

R:Albrecht, J.

submitted to the EMBL Data Library, January 1992

A:Description: Primary structure of the herpesvirus saimiri genome.

A:Reference number: A36806

A:Accession: G36813

A:Molecule type: DNA

A:Residues: 1-407 <ALB>

A:Cross-references: GB:X64346; NID:g60320; PIDN:CAA45696.1; PID:g60384

R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.;

J. Virol. 66, 5047-5058, 1992

A:Title: Primary structure of the herpesvirus saimiri genome.

A:Reference number: A37309; MUID:9233688

A:Contents: annotation; possible protein-coding frames

A:Note: neither amino acid nor nucleotide sequence is given

R:Nicholas, J.; Cameron, K.R.; Honess, R.W.

Nature 355, 362-365, 1992

A:Title: Herpesvirus saimiri encodes homologues of G protein-coupled receptors and cy

A:Reference number: S20243; MUID:92115001

A:Accession: S20244

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-407 <NIC>

A:Cross-references: GB:S76368; NID:g243351; PIDN:AAB21116.1; PID:g243353

C:Genetics:

A:Gene: 73

C:Superfamily: herpesvirus immediate-early protein IE68

C:Keywords: early protein

Query Match 11.3%; Score 88; DB 1; Length 407;

Best Local Similarity 20.0%; Pred. No. 0.085;

Matches 25; Conservative 31; Mismatches 69; Indels 0; Gaps 0;

QY 33 EFRAIDELKQAIEDKEATTATGAASSDALEALADOTDQLOSEAAVVKADNAASDALEA 92

Db 97 EEEAEKEAE 156

A:Accession: S00128  
A:Molecule type: DNA  
A:Residues: 1-593 <OL>  
A:Cross-references: EMBL:X06173; NID:g47084; PIDN:CAA29540.1; PID:g47085  
A:Note: the source is designated as Streptococcus G148  
R:Sjoerling, U.; Falkenberg, C.; Nielsen, E.; Akerstrom, B.; Bjoerck, L.  
J. Immunol. 140, 1595-1599, 1988  
A:Title: Isolation and characterization of a 14-kDa albumin-binding fragment of streptococcal protein G.  
A:Reference number: A27604; MUID:88154455  
A:Accession: A27604  
A:Molecule type: protein  
A:Residues: 62-101 <SUJ>  
R:Guss, B.; Eliasson, M.; Olsson, A.; Uhlen, M.; Frej, A.K.; Jornvall, H.; Flock, J.I.; EMBO J. 5, 1567-1575, 1986  
A:Title: Structure of the IgG-binding regions of streptococcal protein G.  
A:Reference number: A26314; MUID:86300657  
A:Accession: A26314  
A:Molecule type: DNA  
A:Residues: 114-593 <GUS>  
A:Cross-references: GB:X04015; NID:g47071; PIDN:CAA27638.1; PID:g47072  
C:Function:  
A:Description: it is part of the cell wall structure of group G streptococci and is covered by a layer of teichoic acid.  
C:Superfamily: M5 protein  
C:Keywords: duplication; membrane protein  
F:1-33/Domain: signal sequence #status predicted <SIG>  
F:34-593/Product: protein G #status experimental <MAT>  
F:34-116/Domain: alanine-rich <ALA>  
F:117-290/Domain: AB duplication <DUPL>  
F:117-140/Region: A repeat  
F:141-191/Region: B  
F:192-215/Region: A repeat  
F:216-266/Region: B  
F:267-290/Region: A repeat  
F:303-497/Domain: IgG binding <IGB>  
F:303-357/Region: C repeat  
F:358-372/Region: D  
F:373-427/Region: C repeat  
F:428-442/Region: D  
F:443-497/Region: C repeat  
F:498-567/Domain: proline-rich <PRO>  
F:568-593/Domain: carboxyl-terminal <CTD>

Query Match 19.9%; Score 155; DB 2; Length 593;  
Best Local Similarity 56.1%; Pred. No. 5.8e-09;  
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 VDSPIEQPIIPNGTLNLGNAPKALRNEERAIIDELKQAIEDKEATTAIEAASD 60  
DB 34 VDSPIEDTPIIRNGGELTNLGNSETTLALRNEESATDLTAATAAVDTVAANAAGAA 93

QY 61 ALEAA 66  
DB 94 AWEAAA 99

RESULT 3  
A44801  
albumin-binding protein precursor - Streptococcus sp. (fragment)  
C:Species: Streptococcus sp.  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: A44801  
R:Sjoerling, U.  
Infect. Immun. 60, 3601-3608, 1992  
A:Title: Isolation and molecular characterization of a novel albumin-binding protein from Streptococcus sp.  
A:Reference number: A44801; MUID:92363555  
A:Molecule type: DNA; protein  
A:Residues: 1-323 <SJO>  
A:Cross-references: GB:M95520; NID:g153554; PIDN:AAA26847.1; PID:g153555  
A:Experimental source: group G, strain DG12  
A:Note: sequence extracted from NCBI backbone (NCBI:110938, NCBI:110939)  
F:12-323/Product: albumin-binding protein (fragment) #status experimental <MAT>

Query Match 16.2%; Score 126; DB 2; Length 323;  
Best Local Similarity 42.6%; Pred. No. 4.5e-06;  
Matches 26; Conservative 14; Mismatches 21; Indels 0; Gaps 0;

QY 62 LBALADQTDALQSEAAVVKADNAASDALEALADOTDLOQSEAEVQSDNAASDAWEKA 121  
DB 62 LSAQADQIVSAQADNEAITKAEEDSSKAWAADAQANTAKADELAKAEKSSDAWEKA 121

QY 122 A 122  
DB 122 A 122

RESULT 4  
QXBPI1.  
hypothetical protein 401 - phage lambda  
N:Alternate names: orf-401; orf401  
C:Species: phage lambda  
C:Date: 13-Jun-1983 #sequence\_revision 13-Jun-1983 #text\_change 23-Jul-1999  
C:Accession: G43010; D43016; A04389  
R:Daniels, D.  
submitted to the Nucleic Acid Sequence Database, September 1982  
A:Reference number: A94614  
A:Accession: G43010  
A:Molecule type: DNA  
A:Residues: 1-401 <DAN>  
R:Sanger, F.; Coulson, A.R.; Hong, G.F.; Hill, D.F.; Petersen, G.B.  
J. Mol. Biol. 162, 729-773, 1982  
A:Title: Nucleotide sequence of bacteriophage lambda DNA.  
A:Reference number: A92891; MUID:83189071  
A:Accession: D43016  
A:Molecule type: DNA  
A:Residues: 1-401 <SAN>  
A:Cross-references: GB:J02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:g21510  
C:Genetics:  
A:Map position: 40.51-42.99  
C:Superfamily: phage lambda hypothetical protein 401

Query Match 12.7%; Score 99; DB 1; Length 401;  
Best Local Similarity 32.9%; Pred. No. 0.0052;  
Matches 26; Conservative 16; Mismatches 37; Indels 0; Gaps 0;

QY 50 ATTATBAASSDALEALADQTDALQSEAAVVKADNAASDALEALADOTDLOQSEAEVQ 109  
DB 190 AISSSKNAATSAAGAKTSETNAAASQQAATSAATKASEAATGARDAVASKEAKSS 249

QY 110 SDNAASDAWEKAATPTALD 128  
DB 250 ETNASSAGRAASSATAAE 268

RESULT 5  
D90734  
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: D90734  
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: D90734  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-438 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA34267.1; PID:g13360303; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: ECs0844

Query Match 12.4%; Score 96; DB 2; Length 438;





**THIS PAGE BLANK (USPTO)**

```
CC -|- ALTERNATIVE PRODUCTS: THE PRODUCTION OF THE TWO PROTEIN PRODUCTS
CC FROM THIS REGION IS DUE TO PROGRAMED RIBOSOMAL FRAMESHIFTING.
CC -|- SIMILARITY: PARTIAL TO PHAGE T4 DNA POLYMERASE ACCESSORY PROTEIN
CC 44 AND TO HUMAN ACTIVATOR 1, 37 AND 40 KDa SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04487; CAA28174.1; -.
DR EMBL: X04487; CAA28175.1; ALT_INIT.
DR EMBL: X04275; CAA27827.1; -.
DR EMBL: AE000153; AAC73572.1; -.
DR EMBL: U82664; ABA0224.1; -.
DR EMBL: M38777; AAA23457.1; -.
DR PIR: A25549; DJEC3G.
DR ECODBASE; H052.0; 6TH EDITION.
DR ECODBASE; H080.0; 6TH EDITION.
DR Ecodene; EG10245; dnax.
DR InterPro; IPR003959; AAA_subfam.
DR InterPro; IPR000862; RFC.
DR Pfam; PF00004; AAA; 1.
DR Transferrase; DNA-directed DNA polymerase; DNA replication;
DR ATP-binding; Ribosomal frameshift; Complete proteome.
DR CHAIN 1 643 DNA POLYMERASE III SUBUNIT TAU.
DR CHAIN 1 431 DNA POLYMERASE III SUBUNIT GAMMA.
DR NP_BIND 45 52 ATP (POTENTIAL).
DR SEQUENCE 643 AA; 71137 MW; D2028BD99E375150 CRC64;

Query Match 8.8%; Score 68; DB 1; Length 643;
Best Local Similarity 36.5%; Pred. No. 7.3;
Matches 19; Conservative 9; Mismatches 24; Indels 0; Gaps 0;

QY 51 TTAIEAASDDALEALADOTDALQSEAAVVKADNAASDALEALADOTDALQS 102
DB 408 TTSQVLARQQLQVQGATKAKKSEPAATRAPVNNALERLASVTDKRVQA 459

RESULT 40
ID DNAK_CHLMU STANDARD; PRT; 654 AA.
AC P56836.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa
DE protein) (HSP70) (75 kDa membrane protein).
GN DNAK OR TC0675.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=833560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K.,
RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
RA Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G.,
RA Salzberg S.L., Eisen J., Fraser C.M.;
RT Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.*;
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [2]
RP SEQUENCE OF 1-241 FROM N.A.
RC STRAIN=MOPN;
RX MEDLINE=91072247; PubMed=2254267;
RA Engel J.N., Pollack J., Perara E., Ganem D.;
RT "Heat shock response of murine Chlamydia trachomatis.";
```

```
RL J. Bacteriol. 172:6959-6972(1990).
CC -|- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC -|- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING INFECTION.
CC -|- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002336; AAF39496.1; ALT_INIT.
DR EMBL: M62819; AAA23138.1; -.
DR HSSP; P04475; IDGA.
DR TIGR; TC0675; -.
DR InterPro; IPR001023; HSP70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS10336; HSP70_3; FALSE_NEG.
KW Chaperone; ATP-binding; Heat shock; Membrane; Complete proteome.
FT INIT_MET 0 187 0 BY SIMILARITY.
FT CONFLICT 187 187 G -> A (IN REF. 2).
SQ SEQUENCE 654 AA; 70440 MW; EF20C93FE4D3771 CRC64;

Query Match 8.6%; Score 67; DB 1; Length 654;
Best Local Similarity 24.1%; Pred. No. 9.5;
Matches 19; Conservative 19; Mismatches 41; Indels 0; Gaps 0;

QY 39 ELKQAIEDKEATTATAEASSDALEALADOTDALQSEAAVVKADNAASDALEALADOTD 98
DB 567 EKVRAVKREDASTTAIKASDLSARMQKIGEMQAQASANAQGGPNINSEDLKKHSF 626

QY 99 ALQSEAEVQSDNAASDA 117
DB 627 STRPPAGDNSSSTDNIEDA 645

Search completed: October 13, 2002, 02:09:05
Job time : 27.9572 secs
```

```

QY 114 ASDAWKAAATPIALDYVKKTKDKTPVVVKE 142
   :  ||| | : | : | :
Db 449 IRRNAKARTDVELKLSKYOEELAQYKKD 477

RESULT 38
YJFJ_ECOLI
ID YJFJ_ECOLI STANDARD; PRT; 232 AA.
AC P39292;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yjfb precursor.
GN YJFJ OR B4182.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP
SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-95334362; PubMed-7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14003; AAC97078.1; .
DR EMBL; AE000490; AAC77139.1; .
DR EcoGene; EG12485; yjfb.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 18 HYPOHETICAL PROTEIN YJFJ.
FT CHAIN 19 232 HYPOHETICAL PROTEIN YJFJ.
SQ SEQUENCE 232 AA; 25333 MW; 65EB1483D6D29369 CRC64;

Query Match 8.8%; Score 68; DB 1; Length 232;
Best Local Similarity 25.0%; Pred. No. 2.5;
Matches 17; Conservative 14; Mismatches 37; Indels 0; Gaps 0;

QY 39 ELKKQAIEDKEATATATAMSSDALEALADOTDALQSEAAVVKADNAASDALEALADQTD 98
   : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 115 EVSRDGVKAVTATAQRIAFQEOOMEVVKATEAMQRAQQAQVTTSTVCASSSVSTAAESLK 174

QY 99 ALOSEAE 106
   : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 175 RIATROAE 182

RESULT 39
DP3X_ECOLI
ID DP3X_ECOLI STANDARD; PRT; 643 AA.
AC P06710; Q47721;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA polymerase III subunit tau (EC 2.7.7.7) (Contains: DNA polymerase
DE III subunit gamma).
GN DNAX OR DNAX OR DNAX OR B0470.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP
SEQUENCE FROM N.A.

```

```

RC STRAIN-K12 / JM109;
RX MEDLINE-87040775; PubMed-3534795;
RA Flower A.M., McHenry C.S.;
RT "The adjacent dnaX and dnaX genes of Escherichia coli are contained
RT within one continuous open reading frame.";
RL Nucleic Acids Res. 14:8091-8101(1986).
RN [2]
RP
SEQUENCE FROM N.A.
RX MEDLINE-86312915; PubMed-3018672;
RA Yin K.-C., Blinkowa A.L., Walker J.R.;
RT "Nucleotide sequence of the Escherichia coli replication gene dnaX.";
RL Nucleic Acids Res. 14:6541-6549(1986).
RN [3]
RP
SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Coliade-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP
SEQUENCE FROM N.A.
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP
RIBOSOMAL FRAMESHIFTING.
RX MEDLINE-90245360; PubMed-2186364;
RA Blinkowa A.L., Walker J.R.;
RT "Programmed ribosomal frameshifting generates the Escherichia coli
RT DNA polymerase III gamma subunit from within the tau subunit reading
RT frame.";
RL Nucleic Acids Res. 18:1725-1729(1990).
RN [6]
RP
RIBOSOMAL FRAMESHIFTING.
RX MEDLINE-90251627; PubMed-2187190;
RA Flower A.M., McHenry C.S.;
RT "The gamma subunit of DNA polymerase III holoenzyme of Escherichia
RT coli is produced by ribosomal frameshifting.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:3713-3717(1990).
RN [7]
RP
RIBOSOMAL FRAMESHIFTING.
RX MEDLINE-90207226; PubMed-2181440;
RA Tsuchihashi Z., Kornberg A.;
RT "Translational frameshifting generates the gamma subunit of DNA
RT polymerase III holoenzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2516-2520(1990).
RN [8]
RP
REVIEW.
RX MEDLINE-92246902; PubMed-1575709;
RA O'Donnell M.;
RT "Accessory protein function in the DNA polymerase III holoenzyme from
RT E. coli.";
RL Bioessays 14:105-111(1992).
CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC -!- FUNCTION: THE TAU CHAIN SERVES AS A SCAFFOLD TO HELP IN THE
CC DIMERIZATION OF THE CORE COMPLEX.
CC -!- FUNCTION: THE GAMMA CHAIN SEEMS TO INTERACT WITH THE DELTA
CC SUBUNIT TO TRANSFER THE BETA SUBUNIT ON THE DNA.
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + [DNA](N).
CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC DIMERIZATION TO FORM THE POLIIV' COMPLEX. POLIIV' ASSOCIATES WITH
CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC AND CHI) AND WITH THE BETA CHAIN. THE FINAL COMPOSITION OF THE
CC COMPLEX IS: (ALPHA,EPSILON,THETA)[2]-(TAU[2]-(GAMMA,DELTA,DELTA',
CC PSI,CHI)[2]-BETA[4]).

```



[illegible]

